MEASURING BUSINESS RELATED ETHICALITY GLOBALLY: CULTURAL EMIC OR ETIC?

Mary Margaret Rogers, University of New Mexico
Robert A. Peterson, University of Texas, Austin
Gerald Albaum, University of New Mexico

ABSTRACT

When conducting or evaluating cross-cultural/cross-national research studies a critical question must be asked about the measurements: are they culturally an emic or etic? That is, is the research methodology culturally bound or culture free?   The research described in this paper shows how etic and emic properties may be explored by using the Ethicality Scale developed by Albaum and Peterson (2006) and demonstrates that it is at best an imposed etic scale. In doing so, this research confirms Adler's (1991) view that similarity across cultures should be proven rather than assumed. It also raises questions about the frequent assumption of implied or imposed etic validity in cross-cultural/cross-national research.

JEL: M14, M16, M30

KEYWORDS: Equivalency, Ethics, Ethicality Scale, Emic vs. Etic, Cross-cultural Research

INTRODUCTION

Cross-cultural/cross national studies are increasing in number and viewed as providing valuable insight for researchers, classroom teaching, and business practice; yet serious concerns about methodology exist that must be overcome if the studies are to make useful contributions (Hult et al., 2008). Watkins (2010) notes that many researchers make assumptions that "the values measured are 'universal,' exhaustive and applicable to every culture" (p. 702) and often assume validity without establishing it. Thus, when interpreting the results of cross-cultural/cross-national research it is necessary to carefully consider if the measurements are emic or etic.

Among the many issues in cross-cultural/cross-national research (Adler, 1983; Chan & Rossiter, 2003; Sekaran, 1983), perhaps the most important is that of data equivalence (J.W. Berry, 1980; Craig & Douglas, 2000; Kumar, 2000; Mullen, 1995; Salzberger, Sinkovics, & Schlegelmilch, 1999). Studies reviewing international business literature, including international studies in general marketing and buyer behavior have determined that many cross cultural studies do not examine equivalence of data (He, Merz, & Alden, 2008; Hult, et al., 2008; Watkins, 2010). Explanations given for the lack of data equivalence measurement were: data not viewed as conducive to data equivalence measurements; analysis not viewed as necessary; and researchers not familiar with the methodology (He, et al., 2008). Despite numerous calls for improved analysis of data equivalence measures, Hult, et al. (2008) show no statistically significant improvement in data equivalence analysis from 1995-2005, and Watkins (2010) reports such tests are not often presented. Of the researchers in the He, et al. sample not reporting data equivalence measures, 72% indicate that the need for such analysis was not mentioned by either reviewers or editors.

The study reported here involves assessing the validity of the Ethicality Scale, a measurement scale of business-related ethicality developed in the United States by Albaum and Peterson (2006), by measuring business ethics-related attitudes of respondents in 13 countries. The purpose of this assessment is to demonstrate how to test for etic and emic properties by using measures of data equivalence. To do so, the analysis extends guidelines suggested by Hult, et al. (2008) for assessing data equivalence. For the sake of simplicity, the terms "culture" and "nation" are used interchangeably to define the domain of concern.

However, it must be remembered that culture does not always equal nation so that much research that is labeled "cross-cultural" is more appropriately "cross-national."

Below, the concepts of emic and etic research methodology are reviewed followed by a discussion of data equivalence and how equivalence may be measured. The methodology of data collection, measurement, standardization and analysis is then reviewed. Finally, the results of the analysis are discussed and questions for future research are presented.

LITERATURE REVIEW

The Emic/Etic Issue

An aspect of research methodology is an emic when it is culture-bound. That is, it behaves in a specified way in one culture and one culture only. More formally, "Emic validity is established when correct predictions of behavior in a culture are made on the basis of the investigator's understanding of that culture's own conceptual system" (J.W. Berry, 1980 p. 19). When it operates similarly in many cultures, it is considered to be culture-free and is an etic (J.W. Berry, 1980). As noted by Berry (1989), Pike's work on etic and emic behavioral descriptions demonstrates value to the etic approach. An etic study may be a point of entry that provides experience in recognizing similarities and difference, may be a practical way of meeting financial or time limitations in research (Pike, 1967), and may be useful for exploratory research (Cadogan, 2010). Further, Pike notes that etic and emic are not a dichotomy, but two different perspectives that, used jointly, may add richness to the analysis.

However, not all research is exploratory and evidence of cross-cultural validity must be established if thoughtful analysis is to be done in international research. As noted by Vandenberg and Lance (2000) failure to establish between group equivalence can "can render interpretation of between group comparisons on the nonequivalent measures highly suspect" (p. 9). But, by examining methodology in more than one culture, an aspect of validity can be assessed. In doing so, imposed etic validity is established by correctly predicting an outcome in a culture by using a research methodology imported from another culture (J.W. Berry, 1980). That is, "imposed etic validity is established by correctly predicting an outcome in culture B on the basis of a theory, construct, or test imposed from culture A" (Berry, 1980, p. 19).

When emic and imposed etic validities have been "proved," imposed etic validity can be established. Berry (1980, p. 19) argues that this is appropriate for valid cross-cultural comparisons and that imposed etic validity must be based on known validity in two or more cultural systems. Viewed this way, imposed etic validity must be established one culture at a time. In an applied sense, methodologically, this could lead to complications for, say, a business firm that wants to study its corporate reputation in its multiple foreign markets, or even in a small subset of its markets. What if the imposed etic of a methodology held only for some cultures but not others? Then, an overall derived etic validity is unobtainable. That this matter is of current interest to some methodological researchers is illustrated by a recent study of applying a country personality scale developed in a Western country to a Chinese social context to position 11 different countries, including China, on six personality dimensions (d'Astous & Li, 2009). The adapted scale had good psychometric properties in this application.

From a cultural anthropological perspective, the present study can be viewed as a type of hologeistic study of research methodology. A hologeistic or whole earth study is one that uses data from worldwide samples of entire societies or cultures. These samples are intended as representative samples of all known human cultures, or of a defined subset of that universe (Naroll, Michik, & Naroll, 1980). Specifically, it is a holonational study since it uses a sample selected from a population of nation states. According to

Naroll, Michik and Naroll (1980, p. 483), culture presumably varies within the sample as much as usual, other things are presumed equal, and irrelevant factors are presumed to vary randomly.

Equivalence

As discussed above, for imposed etic validity to exist there must be equivalence in effects of research methodology between the nation where the methodology was developed and refined and the nation where it is applied. Equivalence has been defined by Craig and Douglas (2000) 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." Concern with data equivalence requires "taking steps to ensure that any differences found between cultures truly reflect the phenomena of interest, and are not simply a reflection of issues such as scale use tendencies and differences in contract conceptualizations" (Hult, et al., 2008, p. 1028). One approach to equivalence, used in this study, is the psychometric approach in which the characteristics of parameters in measurement models are tested for invariance across countries (van Herk, Poortinga, & Verhallen, 2004). As noted by Steenkamp and Baumgartner (1998), if certain conditions of invariance are satisfied, including configural, metric, scalar, factor covariance, and error variance, then comparisons may be considered valid.

Invariance is established when populations from different cultures that are otherwise identical score identically on a measurement (Schmitt & Kuljanin, 2008). The essence of concern for equivalence in cross-cultural research has been noted in Salzberger (1997), who defined five major dimensions where equivalence is of concern: (1) research methods, (2) research topics, (3) research units, (4) research administration, and (5) data handling.

The dimension of most concern in the present research is that of research methods, an essential aspect of which is concern with equivalence in measurement. This relates to developing so-called scales of measurement for constructs. Researchers have tackled this "problem" (For example, see d'Astous & Li, 2009; Donoho, Herche, & Swenson, 2001; Lages, Silva, Styles, & Pereira, 2009; Mullen, 1995; Parameswaran & Yaprak, 1987; Richins, 1986; Singh, 1995; Steenkamp & Baumgartner, 1998). However, Hult, et al. (2008) recently reviewed 167 studies involving cross-cultural data published in five leading international business journals. A key finding is that researchers report insufficient information regarding data equivalence issues. Following is a brief discussion of the three broad categories of concern for data equivalence, construct, measurement, and data collection equivalence. Interested researchers may see Hult, at al. (2008) for a complete review of technique development.

A demonstration of construct equivalence shows that a concept has the same meaning and functions in the same manner across multiple cultures and entails establishing function, conceptual, and category equivalence before data collection. This means observed behavior must relate to similar problems, or functions (Frijda & Jahoda, 1966); concepts within differing systems of cognition are similar across cultures (McArthur, 2007); and similar groupings, or categorizations, of phenomena occur across cultures (Douglas & Craig, 1983; Kumar, 2000). Pre-data collection construct equivalence may be established through literature reviews, use of existing scales, qualitative fieldwork, and pilot studies. Post-data collection construct equivalence may be established through a variety of statistical analyses including, but not limited to, exploratory and confirmatory factor analysis, Cronbach's alpha, Coefficient Theta, and calculations of composite reliability and average variance extracted, and item total correlation (Anderson & Gerbing, 1988; Bagozzi & Heatherton, 1994; Everitt & Skrondal, 2002; Hult, et al., 2008). Because the focus of the current research was to evaluate an established scale, the results focus on post-data collection construct equivalence issues.

Measurement equivalence, "comparability of the wording, scaling, and scoring of constructs across cultures" (Hult, et al., 2008, p. 1028) has subcategories of calibration, translation, and metric equivalence.

Calibration equivalence entails ensuring measurement units for objective data are comparable; while translation equivalence occurs when identical meaning is delivered across cultures. Finally, metric equivalence exists when "data exhibits similarities of structure within cultures close enough to allow researchers to reasonably assign the majority of the remaining variance to inter-cultural differences" (McArthur, 2007, p. 30). Since the Ethicality Scale under study used perceptual measures, concern for calibration equivalence is not as strong as with research studies using objective measures. However, as discussed below, standardization of data can address calibration issues. Hult, et al. (2008) suggests the use of confirmatory factor analysis for metric equivalence and back translation, translation by committee, or statistical testing for form and meaning equivalence to test translation equivalence.

Finally, data collection equivalence must be established. "Data collection equivalence refers to whether the sources of data, the methods of eliciting data and the resulting samples are comparable across cultures…" (Hult, et al., 2008, p. 1037). This involves sampling frame comparability--a parallel between groups sampled, similar data collection procedures, and a match of sampling methods by using equivalent sampling techniques.

DATA AND METHODOLOGY

The current study was done as part of a larger study measuring the level of business ethicality in multiple countries *(reference to be provided).* The 13 countries selected from the larger data set for this analysis were those where sample size was at least 100.

Sample and Data Collection

To establish data collection equivalence, an attempt was made to obtain reasonably representative samples of undergraduate university business students as research respondents. These students are the future business and political leaders so it is meaningful to assess their ethical beliefs and attitudes. Given the major and widespread ethical and legal lapses that have occurred in the past few years, as illustrated by executives managing such companies as Enron, WorldCom, and Tyco, it is important to know the ethical perspectives of these future leaders. A recent study by Gilley, Robertson, and Mazur (2010) discusses the need for enhancing firm value creation by the development and executive championing of an effective code of ethics. Such "Ethics Code Commitment" affects a broad number of company stakeholders.

A two-stage sampling design was employed in data collection. The first stage consisted of identifying, judgmentally, representative samples of four-year colleges and universities in the countries where data was to be collected. That is, the samples were selected based on the judgment and expertise of the researchers. A judgment sample has potential advantages of developing suitable samples and can provide results as good as probability sampling; and, is the most common approach in these types of studies (Smith & Albaum, 2005). Moreover, the present study was designed to be a broad-based international study, rather than a small, focused study, which might be more accurate, but less generalizable.

The second stage consisted of obtaining a cluster sample of undergraduate business students in each of the stage-one colleges and universities selected. Specifically, to obtain geographically diverse cross-sections of business students, professors in business schools in each country were contacted and asked if their research or teaching assistant would administer the questionnaires to undergraduate business students. Table 1 contains a listing of the countries from which the samples were obtained, the number of colleges and universities sampled in each country, and the number of survey participants from each country. Professors who agreed to participate in the survey were either emailed a copy of the questionnaire or sent blank questionnaires and a preaddressed return envelope. Data collection was accomplished through an in-class setting to control for possible "noise" by having a common data collection environment. The

approach to data collection was chosen to take advantage of the personal relationships that existed between authors and colleagues in the countries where data were obtained. This allowed the investigation to be completed in a reasonable length of time, with a high response rate.

Table 1: Countries Included in Study

Country	Number of Colleges or Universities	Number of Respondents
Brazil	3	131
Canada	3	128
Colombia	3	149
France	2	150
Germany	2	242
Hong Kong	2	113
Morocco	2	109
Norway	2	183
Senegal	3	109
Singapore	2	117
Spain	2	174
Tunisia	3	212
United Kingdom	4	148

This table shows the countries from which data were collected. The second and third columns show the number of universities and number of students in the sample.

The final sample consisted of 1,965 survey participants. In total, there were 911 males and 1,037 females in the obtained sample. The average (mean) age was 23.0 years. The gender distribution and average age of respondents in each country are shown in Table 2. (Because some study respondents did not answer all demographic questions, the demographic group sizes do not sum to the final sample size.) Even though probability sampling was not employed, the samples were deemed to consist of sufficiently broad distributions of undergraduate business students to warrant confidence in the general inferences to be drawn.

Table 2: Demographic Characteristics of Samples

Country	Gender (percent distribution)		N	Mean Age (years)
	Female	Male		
Brazil	39.2%	60.8%	130	21.5
Canada	54.7%	45.3%	128	22.4
Colombia	54.4%	45.6%	147	23.8
France	58.4%	41.6%	149	22.9
Germany	48.1%	51.9%	239	22.8
Hong Kong	53.6%	46.4%	112	20.8
Morocco	52.3%	47.7%	109	21.1
Norway	44.1%	55.9%	179	24.1
Senegal	28.4%	71.6%	109	33.1
Singapore	67.5%	32.5%	117	21.3
Spain	52.7%	47.3%	169	23.1
Tunisia	71.7%	28.3%	212	21.9
United Kingdom	58.8%	41.2%	148	21.2
All Respondents	53.2%	46.8%	1,948	23.0

This table shows demographic characteristics for respondents by country.

Measurement

The Ethicality Scale measure consists of six items scaled as six-category numerical Likert scales, presented from 1 ("strongly agree") to 6 ("strongly disagree"), as shown in Table 3. Only the endpoints of the rating scale were labeled verbally. Thus, the format of the scale items was balanced and did not contain a neutral point. Such a scale assumes that a respondent has an ethics attitude and is able to indicate it. Four of the scale items were reverse-coded (see Albaum & Peterson, 2006) when computing

individual respondent ethicality scores. The range of possible scores was 6 to 36, with the larger the score the greater the degree of business-related ethicality.

In addition to the Ethicality Scale items, the questionnaire included four demographic questions (age, gender, employment status, country of citizenship), and three questions (academic classification, major field of study, and citizenship) used to screen potential survey participants to ensure that the sample was limited to only undergraduate business students from the respective countries.

Table 3: Items Included in Ethicality Scale

1. If a manager in a company is discovered to have engaged in unethical behavior that results primarily in personal gain (rather than corporate gain), he or she should be terminated or fired (reverse coded).
2. If a manager in a company is discovered to have engaged in unethical behavior that results primarily in corporate gain (rather than personal gain), he or she should be terminated or fired (reverse coded).
3. Top business executives should state in no uncertain terms that unethical behaviors in their companies will not be tolerated (reverse coded).
4. It is important that ethical considerations be taken into account when designing company policies (reverse coded).
5. Within a business firm, the ends justify the means.
6. Business behavior that is legal is ethical.

Source: Albaum and Peterson (2006).The items measured by the Ethicality Scale, originally developed by Albaum and Peterson, are shown in the table.

The questionnaire was originally developed in English and pilot-tested on a sample of American business students to obtain a qualitative evaluation of item understandability and assessment ease. It was subsequently translated into Chinese, French, German, and Spanish, usually by professors in the countries where data were collected. Because many of the survey participants were from English-speaking countries, were enrolled in educational institutions where English was the language of instruction, or spoke English or one or more of the languages into which the questionnaire was translated (e.g., Tunisians speak French), there was no need to translate the questionnaire into other languages.

Standardization of Data

In research methodology, the term standardization can be used to refer to the standardization of procedure, interpretations, or scores. Because the research described here involved using a scale across nations/cultures, there is concern with response bias. The most commonly discussed forms of response bias are acquiescence bias, a grouping of responses at one end of the scale, and extreme or modesty response, the selection or rejection of extreme responses (Baumgartner & Steenkamp, 2001; Fischer & Milfont, 2010; Hartgen, Stuart, Walcott, & Clay, 1990). Both forms of bias must be considered before analyzing data and interpreting results, particularly in cross cultural/national research. As noted in Fischer (1990), there is debate about whether differences in response patterns are methodological bias that should be controlled or cultural phenomena that should be studied. However, "cultural tendencies are likely to change the response of participants and make them incomparable across cultural groups" (Fischer, 2004, p. 264). The focus of this research was to examine etic and emic properties of an established scale; thus, we chose to standardize the data before conducting any analysis in to isolate and control for these biases. This allows for a clearer understanding of etic and emic properties.

The standardization of data, often based on mean and dispersion centering, may take several forms: within subject, within group, within culture, or double standardization. Interested readers should see Fischer and Milfont (2010) for a more in depth discussion of various techniques of data standardization. Depending on the focus of the research standardization could occur on numerous levels. We chose to use within-culture means and dispersion indices (Leung & Bond, 1989) where individual observations are standardized using the mean and dispersion, or standard deviation, for all observations from one country.

Thus, the standardized score, y', can be calculated based on the individual observation, x, using the following formula.

$$y' = \frac{x - mean_{culture}}{dispersion_{Culture}} \qquad (1)$$

This type of standardization addresses acquiescent response bias (Fischer & Milfont, 2010). In conjunction reverse coding some of the scale items, it also addresses extreme response bias (Hartgen, et al., 1990).

RESULTS AND DISCUSSION

Initially, two additional, independent samples of undergraduate business students from France and Spain were used to evaluate temporal stability (test-retest reliability) of the Ethicality Scale. The median two-week and one month test-retest correlation measuring stability of the scale was an acceptable 0.62. This is comparable to the finding in the United States reported by Albaum and Peterson (2006).

To establish the existence of construct and measurement equivalence six analyses were done (Table 4). First, the variances of the six Ethicality Scale items were compared across 13 country samples. Although there were some minor differences, item variances were relatively similar, with most standard deviations falling in the range 1.0 to 1.4, suggesting response homogeneity in the context of the business-related ethicality. However, in three countries—Morocco, Senegal, and Tunisia—some standard deviations were as high as 1.7 and 1.8.

Table 4: Construct and Measurement Equivalence Analyses

Type of Equivalence	Method
Construct Equivalence	
Pre-data collection	
Function, conceptual, categorical equivalence	Used existing, validated Ethicality Scale
Post-data collection	
Unidimensionality	Factor Analysis
	Total Item Correlation
Reliability	Coefficient theta
	Comparison of item variance
Construct Validity	Composite reliability
	Average variance extracted
Measurement equivalence	
Calibration equivalence	Perceptual rather than objective measures used
Translation equivalence	Survey translated by native speakers
Metric equivalence	Factor analysis

This table shows how each type of data equivalence, construct and measurement, was tested in the analysis. Subcategories of each type of equivalence are shown for both construct and measurement equivalence.

The second analysis consisted of the six Ethicality Scale items being subjected to separate factor analyses conducted with the widely used Principle Components extraction and Varimax rotation within each of the 13 countries to determine similarities and differences and to see if structures emerged similar to that reported by Albaum and Peterson (2006) for the United States. In that study the first 4 scale items shown on Table 3 were assigned to the Behavioral Ethicality factor, the last two were assigned to the Philosophical Ethicality factor.

Similar to Albaum and Peterson (2006), two factors emerged –Behavior and Philosophy—in all countries but Senegal, where three factors emerged (see Table 5). Where three factors emerged, the third factor consisted of one scale item. The amount of explained variance ranged from 50.75% in Germany to 61.62% in Norway. Again, this is comparable to the 61.10% for the United States reported by Albaum

and Peterson (2006). Table 6 shows, for each country, on which factor each scale item loaded highest. The "rule of assignment" was a factor loading of at least .500 or, if less than this, a large difference in loadings between the two factors.

Table 5: Results of Factor Analyses of Ethicality Scale Items

Country	Number of Factors	Percent Variance Explained
Brazil	2	56.05%
Canada	2	61.62%
Colombia	2	55.01%
France	2	57.78%
Germany	2	50.75%
Hong Kong	2	55.04%
Morocco	2	50.82%
Norway	2	60.69%
Senegal	3	64.74%
Singapore	2	56.56%
Spain	2	54.63%
Tunisia	2	50.91%
United Kingdom	2	59.36%

This table shows the results of a Principle Components with Varimax rotations factor analysis done on the Ethicality Scale, for each country. . The first column shows the number of factors and the second shows the amount of variance explained by the factors.

Table 6: Factor Where Scale Item had Highest Loading, By Country

	Scale Item					
Country	1	2	3	4	5	6
Brazil	B	B	B	n	P	P
Canada	B	B	B	B	P	P
Colombia	B	B	B	B	P	P
France	B	B	B	B	P	P
Germany	B	B	B	B	P	P
Hong Kong	B	B	B	B	P	P
Morocco	P	P	B	B	n	P
Norway	B	B	B	B	P	P
Senegal	P	n	B	B	n	P
Singapore	B	B	B	n	P	P
Spain	B	B	B	P	P	P
Tunisia	B	P	P	n	n	P
United Kingdom	B	B	B	B	P	P

For each county in the study this table shows where each item of the Ethicality Scales loaded on the two main factors, Behavioral Ethicality or Philosophy Ethicality. B= Behavioral Ethicality; P=Philosophy Ethicality; n=neither. See Table 3 for description of scale items.

The third analysis was the calculation of coefficient theta, which is generally viewed as a special case of Cronbach's alpha coefficient. Theta is 'the alpha coefficient for a scale in which the weighting vector has been chosen so as to make alpha a maximum" (Anderson & Gerbing, 1988, p. 61). It is useful for analyzing a multi-dimensional scale with heterogeneous relationships among the scale items such as the Ethicality Scale (Dillon & Goldstein, 1984). The results are shown in Table 7. The majority are lower than Nunnally and Bernstein's (1994) suggested norm of 0.7 for coefficient alpha. Yet, the number of items in the scale is not large. Some researchers have shown that small numbers of items have small alphas. Or, to put it another way, as the number of items increase, Coefficient alpha also increases (Cortina, 1993; Duhachek, 2004; Spector, 1992). However, 2 of the 13 countries have thetas exceeding the norm for alpha, and another 8 countries are within 0.1 of the expected norm.

Next, we calculated both composite reliability and average variance extracted (AVE) for the overall scale and each factor as suggested by (Anderson & Gerbing, 1988). Composite reliability measures how the underlying factors contribute to the measurement of the construct. Average AVE assesses the amount of common variance among the underlying factors explained by the construct. An AVE of 0.5 or higher is generally viewed as an indication of construct validity (Bagozzi & Heatherton, 1994; Dillon & Goldstein,

1984). As shown in Table 8, Morocco, Senegal, and Tunisia all have composite reliabilities of less than 0.7 and an AVE of less than 0.5 indicating the Ethicality Scale should not be viewed as etic in those countries. However, all other countries, except Spain have composite reliabilities for both the overall scale and each factor of above or very close to 0.7. In addition each of these countries and Spain has AVE's above 0.5 providing evidence that the scale may be used and assumed etic except in Morocco, Senegal, and Tunisia.

Table 7: Coefficient Theta

Country	Theta	N
Brazil	0.677	131
Canada	0.682	128
Colombia	0.650	149
France	0.673	150
Germany	0.601	242
Hong Kong	0.600	113
Morocco	0.539	109
Norway	0.737	183
Senegal	0.485	109
Singapore	0.637	117
Spain	0.628	174
Tunisia	0.589	212
United Kingdom	0.727	148

This table shows the result of a coefficient theta calculation, a special case of Cronbach's alpha, and a means of testing for construct reliability.

Table 8: Composite Reliabilities and Average Variance Extracted

Country	Overall Composite Reliability	AVE	Behavior Ethicality Composite Reliability	AVE	Philosophy Ethicality Composite Reliability	AVE	N
Brazil	0.738	0.664	0.719	0.620	0.774	0.745	131
Canada	0.785	0.770	0.746	0.683	0.861	0.906	128
Colombia	0.729	0.644	0.726	0.638	0.735	0.657	149
France	0.760	0.715	0.742	0.675	0.796	0.791	150
Germany	0.711	0.603	0.699	0.575	0.736	0.659	242
Hong Kong	0.756	0.706	0.735	0.657	0.798	0.796	113
Morocco	0.643	0.448	0.584	0.329	0.746	0.683	109
Norway	0.758	0.711	0.725	0.636	0.822	0.842	183
Senegal	0.419	0.115	0.536	0.250	-0.203	0.007	109
Singapore	0.733	0.654	0.705	0.587	0.789	0.777	117
Spain	0.703	0.584	0.677	0.523	0.753	0.698	174
Tunisia	0.565	0.297	0.578	0.320	0.533	0.245	212
United Kingdom	0.771	0.740	0.746	0.683	0.821	0.840	148

This table show the results of composite reliability and average variance extracted for the entire Ethicality Scale (overall), and for those items loading on the Behavior Ethicality factor or the Philosophy Ethicality factor. These measurements are used to test for construct validity.

Our sixth analysis was a calculation of item-total correlation. This analysis is used to evaluate if a single measure, in this case ethicality, can be used for comparison across a population. If correlations are below 0.2 or 0.3 it is generally accepted that the item doesn't fit well with other measures and should be deleted(Everitt & Skrondal, 2002; Field, 2005) . As shown in Table 9, all scale items except item 5, 'Within a business firm, the ends justify the means' had correlations above 0.2 and most were above 0.3. However, for item 5, 9 of the 13 countries in the sample had total-item correlations below 0.2. This indicates this item may not be an acceptable scale measure.

Perusal of the literature of cross-cultural/national studies, and studies within a culture/nation, relevant to many disciplines suggests imposed etic validity is assumed, without testing, as research methodologies developed and refined in one culture/nation are applied in other cultures/nations. In short, there are few attempts made to assess, empirically, etic properties even though it may be reasonable to assume that some aspects of methodology may better apply when treated as an emic. When a research methodology is "applied" there is an interaction between that methodology and the research respondents or subjects.

Since it is well known that people in different cultures/nations may differ in such basic characteristics as values (Hofstede, 2001; Kahle, Rose, & Shoham, 2000), it would be prudent to question at the outset the assumption of imposed etic validity for most aspects of method. This view is consistent with that proposed by Adler (1991), "assume difference until similarity is proven" (p. 67), and "it remains best to resist the temptation of assuming that any particular theory applies everywhere" (Adler, 2002, p. 165).

Table 9: Item Total Correlations

Country	Scale Item					
	1	**2**	**3**	**4**	**5**	**6**
Brazil	0.672	0.684	0.522	0.463	0.146	0.276
Canada	0.605	0.686	0.558	0.418	0.400	0.344
Colombia	0.636	0.687	0.628	0.485	0.157	0.567
France	0.624	0.694	0.593	0.449	0.168	0.453
Germany	0.650	0.514	0.580	0.491	0.174	0.382
Hong Kong	0.619	0.580	0.603	0.632	0.463	0.414
Morocco	0.609	0.596	0.477	0.505	0.298	0.584
Norway	0.525	0.599	0.583	0.407	0.139	0.375
Senegal	0.470	0.287	0.549	0.562	0.131	0.533
Singapore	0.576	0.594	0.633	0.430	0.262	0.339
Spain	0.646	0.543	0.593	0.367	0.100	0.465
Tunisia	0.527	0.559	0.678	0.408	0.060	0.483
United Kingdom	0.674	0.633	0.700	0.438	0.075	0.213

This table shows the results of the Item Total correlations used to test for unidimensionality. Each column corresponds to an item on the Ethicality Scale shown in Table 3. See Table 3 for description of scale items.

The results of the present study seem to support this notion. There is variation in coefficient theta scores indicating internal consistency reliability is not universal by any means. In contrast, the results of the factor analyses, composite reliability, and AVE generally support that the Ethicality Scale can be viewed as an etic, in some, but not all, countries. This is further supported by the item-total correlations that showed one of the scale items assigned to the Philosophical Ethicality factor not to warrant inclusion in the scale.

Thus, the underlying structure of this scale, and any measurement instrument, should be empirically examined in any cross-cultural or cross-national study. For example, take the case of the country Senegal. Coefficient theta of the overall Ethicality Scale was the lowest at 0.485. The factor analysis yielded three factors, not two. In forcing two factors, the amount of variance explained decreased from 64.74% to 47%, a decrease of more than 25 percent. Finally, composite reliability and AVE were also low at 0.419 and 0.115, respectively. Here, as well as in Tunisia and Morocco, it is only reasonable to consider the scale to be emic. However, the results also indicate scale items, with the exception of item 5, might be considered etic in the other countries. However, great care should be used if including the fifth scale item, 'Within a business firm, the ends justify the means'. These findings also correspond to Pike's (1967) view that the concepts of etic and emic are not dichotomous. Rather they may be more usefully interpreted as endpoints on a scale.

CONCLUDING COMMENTS

The goal of the work discussed here was to demonstrate how to examine etic and emic properties in cross-cultural research. Data was collected using the Ethicality Scale developed by Albaum and Peterson (2006). Analyses were conducted to test for construct and measurement equivalence. The results indicate the Ethicality Scale may be used in many but not all countries and confirmed the need to carefully test any measurement scale for etic/emic properties.

The present study clearly has some limitations that future studies would be well advised to avoid, if possible. First, the sizes of the samples within each country varied widely. Such variation makes

generalization difficult even with standardization of the data. Future research should use samples of approximately equal size, preferably larger samples. Second, it is difficult to make generalizations about any other scale of measurement on the basis of the findings from the present study. The present study does suggest, however, that researchers wanting to use a measurement scale that has not been shown to have imposed etic validity in a country different from the one where the scale was developed should at the very least do a pretest or exploratory study to assess its potential psychometric properties.

These results lead to broader questions: 1) Should a researcher have to test application of all aspects of a research methodology to be used for imposed etic validity properties? 2) Can researchers trust the results of other methodological studies, regardless of the culture being investigated? 3) Is it practical (i.e., economically feasible) to test etic properties for all cultures (nations) or can similarity be assumed, at least for nations having low psychic/psychological distance (Albaum & Duerr, 2011), as measured by culture, stage, of economic development, history, etc., between them? In short, can cultures (nations) be grouped in some meaningful way to ease the difficulty of testing for etic properties? 4) If one assumes there always will be some differences, is there an acceptable level of difference in method effects, and how much tolerance can academic and practitioner researchers accept?

Imposed etic validity can be empirically assessed, as has been done with the Ethicality Scale. As shown, the analyses to test for etic and emic properties are not difficult and should be conducted for all cross-cultural research. However, or, perhaps unfortunately, they may not lead to the clarity of results desired by researchers. While the investigated etic/emic properties of the Ethicality Scale show the scale cannot be assumed to be etic across all cultures/nations, it may be safe for some. But clearly, caution should be used when applying the scale all individual or multiple culture studies.

REFERENCES

Adler, N. J. (1983). A typology of management studies involving culture. *Journal of International Business Studies, 14*(2), 29-47.

Adler, N. J. (1991). *International dimensions of organizational behavior*. Boston: Kent.

Adler, N. J. (2002). *International dimensions of organizational behavior* (Fourth ed.). Cincinnati, OH: South-Western.

Albaum, G., & Duerr, E. (2011). *International marketing and export management* (7th ed.): Prentice Hall.

Albaum, G., & Peterson, R. A. (2006). Ethical attitudes of future business leaders: do they vary by gender and religiosity? *Business & Society, 45*(3), 300-321. doi: 10.1177/0007650306289388

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411-423. doi: 10.1037/0033-2909.103.3.411

Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling: A Multidisciplinary Journal, 1*(1), 35-67.

Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 143-156.

Berry, J. W. (1980). Introduction to methodology. *Handbook of cross-cultural psychology, 2*, 1-28.

Berry, J. W. (1989). Imposed Etics-Emics-Derived Etics: The Operationalization of a Compelling Idea. *International Journal of Psychology, 24*(6), 721.

Cadogan, J. (2010). Comparative, cross-cultural, and cross-national research: A comment on good and bad practice. *International Marketing Review, 27*(6), 601-605. doi: 10.1108/02651331011088245

Chan, A., & Rossiter, J. (2003). *Measurement issues in cross cultural values research*.

Cortina, J. (1993). What is Coefficient Alpha? A Test of Theory and Applications. *Journal of Applied Psychology, 78*(1), 98–104.

Craig, C., & Douglas, S. (2000). *International marketing research*: Wiley New York.

d'Astous, A., & Li, D. (2009). Perceptions of countries based on personality traits: a study in China. *Asia Pacific Journal of Marketing and Logistics, 21*(4), 475-488.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis* (Vol. 394): Wiley New York.

Donoho, C. L., Herche, J., & Swenson, M. J. (2001). *Assessing the transportability of measures across cultural boundaries: a personal selling context.* Paper presented at the Cross Cultural Research Conference, Turtle Bay, Hawaii.

Douglas, S. P., & Craig, C. (1983). *International marketing research*: Prentice-Hall (Englewood Cliffs, NJ).

Duhachek, A. (2004). Alpha's Standard Error (ASE): An Accurate and Precise Confidence Interval Estimate. *Journal of Applied Psychology, 89*(5), 792.

Everitt, B., & Skrondal, A. (2002). *The Cambridge dictionary of statistics* (Vol. 4): Cambridge University Press Cambridge.

Field, A. (2005). Exploratory factor analysis. *Discovering statistics using SPSS, 2*, 619-680.

Fischer, R., & Milfont, T. L. (2010). Standardization in psychological research. *International Journal of Psychological Research, 3*(1), 88-96.

Frijda, N., & Jahoda, G. (1966). On the scope and methods of cross-cultural research. *International Journal of Psychology, 1*(2), 109-127.

Gilley, K. M., Robertson, C. J., & Mazur, T. C. (2010). The bottom-line benefits of ethics code commitment. *Business Horizons, 53*(1), 31-37.

Hartgen, D. T., Stuart, A. W., Walcott, W. A., & Clay, J. (1990). ROLE OF TRANSPORTATION IN MANUFACTURERS'SATISFACTION WITH LOCATIONS. *Transportation Research Record*(1274).

He, Y., Merz, M. A., & Alden, D. L. (2008). Diffusion of measurement invariance assessment in cross-national empirical marketing research: perspectives from the literature and a survey of researchers. *Journal of International Marketing, 16*(2), 64-83.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*: Sage Pubns.

Hult, G., Ketchen, D., Griffith, D., Finnegan, C., Gonzalez-Padron, T., Harmancioglu, N., . . . Cavusgil, S. (2008). Data equivalence in cross-cultural international business research: assessment and guidelines. *Journal of International Business Studies, 39*(6), 1027-1044.

Kahle, L., Rose, G., & Shoham, A. (2000). Findings of LOV throughout the world, and other evidence of cross-national consumer psychographics: Introduction. *Journal of Euromarketing, 8*(1), 1-13.

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

Lages, L. F., Silva, G., Styles, C., & Pereira, Z. L. (2009). The NEP Scale: A measure of network export performance. *International Business Review, 18*(4), 344-356. doi: 10.1016/j.ibusrev.2009.04.002

Leung, K., & Bond, M. H. (1989). On the empirical identification of dimensions for cross-cultural comparisons. *Journal of Cross-Cultural Psychology, 20*(2), 133-151.

McArthur, D. N. (2007). Construct equivalence in international business research: the first and the last of it. *Journal of Business Inquiry, 6*(1), 28-38.

Mullen, M. R. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies, 26*(3), 573-596.

Naroll, R., Michik, G., & Naroll, F. (1980). Holocultural research methods. *Handbook of cross-cultural psychology, 2*, 479-521.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Parameswaran, R., & Yaprak, A. (1987). A cross-national comparison of consumer research measures. *Journal of International Business Studies, 18*(1), 35-49.

Pike, K. L. (1967). *Language in Relations to a Unified Theory of the Structure of Human Behavior* (2nd, revisied edition ed.). The Hague, The Netherlands: Mouton & Co.

Richins, M. L. (1986). *Adapting Psychometric Measures for Use in Cross-cultural Consumer research.* Paper presented at the Annual Conference of the Institute for Decisions Sciences.

Salzberger, T., Sinkovics, R., & Schlegelmilch, B. (1999). Data equivalence in cross-cultural research: a comparison of classical test theory and latent trait theory based approaches. *Australasian Marketing Journal, 7*(2), 23-38.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*(4), 210-222.

Sekaran, U. (1983). Methodological and theoretical issues and advancements in cross-cultural research. *Journal of International Business Studies, 14*(2), 61-73.

Singh, J. (1995). Measurement issues in cross-national research. *Journal of International Business Studies, 26*(3), 597-619.

Smith, S. M., & Albaum, G. (2005). *Fundamentals of Marketing Research*. Thousand Oaks, CA: Sage Publications.

Spector, P. (1992). *Summated rating scale construction: An introduction*: Sage Pubns.

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research, 25*(1), 78-90.

van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response Styles in Rating Scales: Evidence of Method Bias in Data From Six EU Countries. *Journal of Cross-Cultural Psychology, 35*(3), 346-360. doi: 10.1177/0022022104264126

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods, 3*(1), 4.

Watkins, L. (2010). The cross-cultural appropriateness of survey-based value(s) research: A review of methodological issues and suggestion of alternative methodology. *International Marketing Review, 27*(6), 694-716. doi: 10.1108/02651331011088290

BIOGRAPHY

Mary Margaret Rogers is an Associate Professor at the Anderson Schools of Management, The University of New Mexico. He research interests include sustainability and resilience of supply chains and business ethics.

Robert A. Peterson is associate vice president for research at The University of Texas at Austin and holds the Stuart Chair in the McCombs School of Business there. His research interests range from marketing and advertising strategy to behavioral and methodological topics.

Gerald Albaum is Research Professor at the Anderson Schools of Management, The University of New Mexico. He is also Senior Research Fellow at the IC2 Institute at the University of Texas at Austin. His research interests range from research methodology to business ethics.