International Business

Research, Teaching and Practice
The journal of the AIB-SE

International Business: Research Teaching and Practice
2009 3(1)

OVERCOMING COMMON PITFALLS IN CROSS CULTURAL MANAGEMENT RESEARCH

Daniel Rottig*
Florida Gulf Coast University
10501 FGCU Blvd S., Fort Myers, FL 33965

Although cross-cultural research has become a major field of academic study in management, researchers are still struggling with a number of methodological problems. This paper addresses three key problems that frequently occur in empirical, cross-cultural management research: the level of analysis pitfall, the equivalence/invariance pitfall, and the sampling pitfall. These pitfalls are discussed in detail and strategies are presented to overcome these problems.

INTRODUCTION

Culture is an important construct in international management, and has become a major topic of academic study. However, there exist a number of methodological problems with cross-cultural research. The three most frequently occurring methodological problems are the (1) level of analysis pitfall, (2) equivalence/invariance pitfall, and (3) sampling pitfall.

* Telephone: (239) 590-1247
Email: DRottig@fgcu.edu

PROBLEM I - THE LEVEL OF ANALYSIS PITFALL

Hofstede (1980, 2001), in his now ubiquitous study on Culture's Consequences, notes that "[i]n studying "values" we compare individuals; in studying "culture" we compare societies" (2001: 15). With this notion, Hofstede addresses the level of analysis issue, a topic that has been introduced as early as 1939 by Thorndike. The level of analysis issue has been emphasized extensively in theoretical accounts (e.g. Castro, 2002, Dansereau, Yammarino and Kohles, 1999, Glick, 1985, Hofmann, 1997, Hofstede, Bond and Luk, 1993, House, Rousseau and Thomas-Hunt, 1995, Klein, Tosi and Cannella Jr., 1999, Klein, Dansereau and Hall, 1994, Lenartowicz and Roth, 1999, Leung and Bond, 1989, Rousseau, 1985, Schriesheim, 1995), but unfortunately is often either not recognized, or not treated carefully in cross-cultural studies (e.g. Hofstede, 1980, 2001, Hofstede, et al., 1993).

The level of analysis pitfall "occurs when conclusions applying to one level have to be drawn from data only available at another" (Hofstede, et al., 1993: 484). In other words, researchers often erroneously assume that relationships between variables, such as measured by a correlation coefficient, are equal across levels of analysis (Glick, 1985, Hofstede, 2001). This pitfall is commonly referred to as cross-level fallacy (Rousseau, 1985), and may occur in the two types of ecological fallacy and reverse ecological fallacy (Robinson, 1950, Thorndike, 1939). The former suggests that a relationship at a higher level of analysis (e.g. societal/national level, regional level) may not apply at a lower level of analysis (e.g. individual level, organizational level); the latter describes the exact opposite – that a relationship at a lower level of analysis may not apply at a higher level of analysis.

Ecological/Reverse Ecological Fallacy

An often committed ecological fallacy in cross-cultural management research is to use cultural dimensions that were conceptually and methodologically developed at the national/societal level of analysis and interpret them as if they were directly applicable to individuals within the respective national cultures. For example, the finding that a particular culture, A, is relatively masculine when compared to another culture, B, does not imply that every individual within culture A is characterized by masculinity-oriented values, such as high levels of assertiveness, need for achievement and competitiveness. Instead, such a finding implies that a country ranking high on the masculinity dimension has a population of inhabitants that ***on average*** shows a higher within-society correlation for than a country with culture B for that dimension. In fact, when Hofstede developed his original four national cultural dimensions in the 1970s, the individual level data structure was not helpful to create distinct cultural categories. This analysis result led Hofstede to conceptualize his cultural dimensions on the societal level of analysis and report his measures for the societal level metric structure (Hofstede, 1980).

The opposite of an ecological fallacy is the reverse ecological fallacy, which can be equally as misleading as the ecological fallacy. Here, researchers do not solely commit the pitfall of inadequate data treatment, but also are trapped in the snare of basing their work on an inadequate research paradigm. As noted by Hofstede, "(c)ultures are not king-size individuals: They are wholes, and their internal logic cannot be understood in the terms used for the personality dynamics of individuals. Eco-logic differs from individual logic" (2001: 17). In other words, researchers who treat cultures as if they were individuals commit a reverse ecological fallacy, and therefore likely arrive at misleading or erroneous conclusions.

Strategies to Overcome Cross-level Fallacies

The aforementioned explications, however, do not imply that multilevel research should not be conducted or is not desirable. On the contrary, many authors strongly advocate multilevel research (e.g. House, et al., 1995) and discuss methods that facilitate the study of a dataset at more than one level of analysis (e.g. Castro, 2002, Hofmann, 1997). Relationships at different levels of analysis have their own logic, function, and meaning, requiring researchers to be aware of the conceptual and methodological implications of analyzing a data set at different levels of analysis.

Select the Appropriate Type of Analysis for Cross-cultural Comparisons. One strategy to overcome the level of analysis pitfall is to ensure selection of the appropriate statistical technique. A data set comprising 1,000 individuals from 10 cultures, for example, may be examined in several ways depending on the focal level of analysis. These alternative types of analysis include:

1) Correlation analysis across all 1,000 individuals, regardless of their cultural affiliation, which is referred to as "pancultural" (Leung and Bond, 1989: 135) or "global" correlation (Hofstede, 2001: 16);
2) Separate correlation analysis of the focal variables within each culture, resulting in 10 correlations (one for each culture) between each pair of variables, referred to as "intracultural" (Leung and Bond, 1989: 135), or "within-society" correlations (Hofstede, 2001: 16);
3) Computation of the means of the focal variables for each of the 10 cultures and correlation analysis of these mean scores across the 10 cultures, referred to as "ecological" (Robinson, 1950: 351), "cross-cultural" (Leung and Bond, 1989: 135), or "between-society" correlation (Hofstede, 2001: 16);

4) Correlation analysis across all 1,000 individuals, regardless of their cultural affiliation, by using pooled data after elimination of the culture-level effect, referred to as individual level analysis (Hofstede, et al., 1993: 487, Leung et al., 2005: 144).

It is important to emphasize that although the aforementioned correlations are based on the same dataset, in most case these correlations will not be equivalent. In fact, the latter two ways of analyzing a dataset (ecological and individual level analyses) are entirely unrelated and may lead to correlation coefficients that may not only differ in strength, but also in direction. Further, Leung and Bond (1989) found that the results of pancultural analysis (analysis type number 1) resemble the results of ecological analysis (type 3) more than they resemble the results of individual analysis (type 4).

The aforementioned four types of cross-cultural comparisons also have implications in factor analysis. If a researcher attempts to reveal a structure underlying a set of items using the dataset mentioned above (a sample size of 1,000 individuals from 10 cultures), the following factor analyses should be performed for pancultural, intracultural, ecological, and individual correlations, respectively:

1) Inclusion of all 1,000 subjects into a factor analysis of all items (i.e. one factor analysis using raw scores).
2) Separate factor analyses of all items for each of the 10 cultures and inclusion of only the subjects who are affiliated with the respective culture (i.e. 10 separate factor analyses using raw scores).
3) Factor analysis of the means of the focal variables across the 10 cultures (i.e. one factor analysis using means).
4) Iinclusion of all 1,000 subjects into factor analysis of all items, after elimination of the culture-level effect (i.e. one factor analysis using adjusted raw scores).

Like the previously described relationships between correlations, the results of these factor analyses, although based on the same dataset in most cases will not be equivalent (Hofstede, et al., 1993, Leung and Bond, 1989). Being aware of the aforementioned distinctions between types of cross-cultural comparisons decreases the probability that researchers will commit a cross-level fallacy, and therefore contributes to overcoming this pitfall.

Remedies for Reverse Ecological Fallacy In comparing the two types of cross-level fallacies (i.e. ecological and reverse ecological fallacy), Hofstede et al. note that the reverse ecological fallacy "has been much less often recognized, and is thus more frequently committed" (1993: 485). The authors assert that this is based on the impact of culture on research in general, and the individualistic bias of researchers in particular. Because Western countries (such as the United States, Canada, Western Europe, and Australia) are primarily characterized by individualistic values, aspects of social sciences that largely deal with individual-level data (such as psychology and social psychology) are emphasized in these countries compared to disciplines that concern ecological-level data (such as sociology, political science and anthropology). Whereas researchers in the latter fields of social sciences are more prone to commit an ecological fallacy, scholars belonging to the former disciplines are more tempted to be trapped in a reverse ecological fallacy.

A first step for researchers to overcome this pitfall is to become explicitly aware that analysis results based on individual-level data may not apply at a higher level of analysis. In order to draw meaningful conclusions and derive plausible interpretations concerning relationships at a higher level of analysis, researchers need to aggregate their individual-level data to the respective higher level. Whether aggregating individual-level measures to a higher level of analysis is significantly justified needs to be determined.

An initial idea about whether such an aggregation will be meaningful can be obtained by examining the statistical relevance of higher-level variables. In the aforementioned database with 1,000 individuals from 10 countries, researchers may want to see whether the country affiliation of the respondents has a significant effect on the analysis results. This could be done by running an analysis of covariance that specifies the country variable as a covariate (i.e. a variable with 10 categories, one for each country) and examining its statistical significance. More sophisticated procedures concern the use of multilevel empirical analyses, such as intraclass correlation coefficients, hierarchical linear modeling, rwg(j), within- and between-analysis, and random group sampling (for a more detailed discussion about these procedures, see e.g. Castro, 2002).

Aggregation Procedures. Once aggregated, researchers need to understand that the aggregated scores now describe relationships at a higher level of analysis, and no longer explain relationships at the individual level (Hofstede, et al., 1993). Only if aggregated scores are used for the statistical analysis can meaningful conclusions about a respective higher level of analysis be drawn. In deciding how to aggregate measures, researchers may (1) create individual level scales and subsequently aggregate them to a higher level of analysis, which often concerns gathering data on individual variables (such as personal values) that necessarily need to be collected at the individual level, (2) aggregation of items and subsequent creation of aggregate scales, which is typically used where primary interest lies in societal values (societal norms, institutional practices etc.) and therefore conceptually treats individual-level items as indicators of societal values, and (3) a combination of the aforementioned two approaches, which involves the creation of individual-level scales, aggregation to a higher level, and subsequent creation of aggregate scales (Peterson, 2005). The latter approach provides a stronger basis for aggregate measures design, but also requires a larger number of items and societies.

A Framework for Cross-cultural Assessment

In an attempt to refocus cross-cultural research from the predominant country-based definition of culture to a cultural grouping(s) perspective, Lenartowicz and Roth (1999) developed a framework for cultural assessment. In so doing, these authors introduced an additional level of analysis, the cultural grouping (see also, Lenartowicz and Roth, 2001). The framework suggests that, depending on the level of analysis and focus of observation, cross-cultural management researchers may use either comparative (individuals grouped by cultures and involving a few cultures), unicultural (individuals from one culture), cross-cultural (involving many cultures), or intercultural (individuals from a few cultures) analysis. The framework is presented in the context of a general pattern of:

1) Validation of establishing the unit of analysis,
2) Screening of selected subjects,
3) Verification of homogeneity of the cultural groups,
4) Use of appropriate measures, and
5) Validation of sample representativeness.

The previous approaches apply to other than explicitly mentioned levels of analysis. In other words, whereas, the previous discussion concentrated on individual versus societal/national level of analysis, the logic and explications provided are equally relevant for the group, organizational subsystem, organizational, industry, organizational population, organizational field, intra-country regional (sub-cultural), regional, country-cluster, or global system levels of analysis.

Problem II - The Equivalence/Invariance Pitfall

Measurement non-invariance (sometimes referred to as construct bias) means that a set of measures (items) fails to assess the same latent variables in different groups (Kline, 2005). If a measurement scale pre-validated in one cultural setting does not lead to the same factor structure in another culture (due to respondents from different cultures associating different items with different constructs) then the measurement scale is non-invariant or non-equivalent across cultures (Cheung & Rensvold, 1999). This pitfall of measurement equivalence/ invariance constitutes a major problem faced by researchers, and can be quite detrimental to their analysis results and interpretations (Cheung and Rensvold, 1999, Steenkamp and Baumgartner, 1998).

Oftentimes, scholars frequently and implicitly assume measurement equivalence, but fail to examine whether their measures are, in fact, invariant (Steenkamp & Baumgartner, 1998). Instead, researchers regularly focus solely on establishing general psychometric properties of their measurement scales (such as validity and reliability), but do not test for (a) conceptual equivalence of the underlying theoretical variables, (b) whether there are equivalent associations between operationalizations and the theoretical variables, and (c) the extent to which responses across cultural groups are influenced to the same degree by the same underlying factors (Kline, 2005). Conclusions drawn from findings based on invariant measures "are at best ambiguous and at worst erroneous" (Steenkamp and Baumgartner, 1998: 78).

Strategies to Overcome the Measurement Non-Invariance Pitfall

In order to overcome the non-invariance problem, researchers need to explicitly recognize that testing for measurement invariance/equivalence is of paramount importance Both traditional and contemporary approaches are available.

Traditional Approaches. The issue of invariance has been identified by other names, such as, for example, problem of response sets (Stening and Everett, 1984). A relationship between two unrelated variables may occur due to cultural differences in respondents' willingness to use the extreme ends of measurement scales, tendency towards acquiescence (Leung and Bond, 1989), predisposition to other types of common method bias such as social desirability, consistency or likeability (Podasakoff, MacKenzie, Lee and Podsakoff, 2003), as well as types of systematic biases distinctive of individuals with certain cultural affiliations (Steenkamp and Baumgartner, 1998). Traditional methods of dealing with the problem of response sets involve either standardizing two unrelated variables within each culture prior to running statistical analyses, or within unit-of-analysis standardization (Leung and Bond, 1989). Given that the former approach is plagued by the problem of eliminating true differences at the cross-cultural-level, the latter method is preferred (Leung and Bond, 1989). Within unit-of-analysis standardization involves standardization of responses to a set of items within each unit ("unit" referring to individual respondents at the individual level of analysis and to cultures at the ecological level of analysis). More detailed illustrations of these traditional approaches can be found in Hofstede (1980, 2001) as well as in the Chinese Culture Connection (1987).

Contemporary Approaches. Contemporary methods to overcome the problem of measurement invariance involve structural equation modeling procedures, such as multigroup path analysis and multigroup confirmatory factor analysis (Kline, 2005). The former is used when constructs are assessed by one-item measures, and addresses the problem of whether estimates of model parameters vary across cultural groups and tests whether cultural group membership moderates the relationship of interest (Kline, 2005). A more sophisticated approach is to impose equality constraints across the cultural groups and examine the differences across groups by comparing analysis results and modification indices.

Another contemporary method to test for measurement invariance is Multiple Group Confirmatory Factor Analysis (CFA). In situations where latent constructs are measured by a set of indicators (items) across cultural groups, multiple group CFA can be used to test for invariance. This analysis tool addresses the question of whether the focal set of indicators assesses the same latent variables across cultural groups. This procedure involves the specification of two CFA models, one with equality constraints imposed across cultural groups, and another with parameters freely estimated (Kline, 2005). These two models are compared based on relative fit by using a chi-square difference test (Kline, 2005) or by examining the change in other goodness-of-fit indices as identified by Cheung and Rensvold (2002). The equality constraint imposed between multiple group models depends on the focus of the study and the goals of the researcher, and the minimum level of invariance that needs to be present differs across these foci and goals (Steenkamp and Baumgartner, 1998).

Vandenberg and Lance (2000) propose a set of eight measurement invariance tests, arranged in a framework of sequential steps using multigroup CFA that belongs to the family of structural equation modeling-based analysis tools. In fact, although researchers still debate about different types of invariance, "there is general agreement that the multigroup confirmatory analysis model (Jöreskog, 1971) represents the most powerful and versatile approach to testing for cross-national measurement invariance" (Steenkamp and Baumgartner, 1998: 78). Table 1 provides an overview of these eight steps. The following discussion assumes a basic understanding of structural equation modeling (SEM) summarized in basic texts on SEM such as Raykov and Marcoulides (2000), Kline (2005), and Byrne (1998).

1) *Test for Invariant Covariance Matrices:* This initial test examines whether covariance matrices are invariant across cultural groups. With two cultural groups, for example, researchers need to examine the null hypothesis that both covariance matrices are equivalent/invariant. This evaluation is based on the common SEM procedure of assessing goodness of fit by examining the likelihood ratio chi-square ($\chi 2$) statistic, $\chi 2$ by degrees of freedom ratio, as well as other fit indices, such as Goodness-of-fit index (GFI) (e.g. Medsker, Williams and Holahan, 1994, Mulaik, et al., 1989), Steiger and Lind's (1980) Root Mean Square

Table 1. Testing for Equivalence/Invariance Using Multigroup Confirmatory Factor Analysis (CFA)

Test 1 Basic Invariance	Test 2 Invariance of Psychometric Properties	Test 3 Structural Invariance
Step 1 Test for Invariant Covariance Matrices	*Step 2* Test for Configural Invariance	*Step 6* Test for Construct Variance Invariance
	Step 3 Test for Metric Invariance	*Step 7* Test for Construct Covariance Invariance
	Step 4 Test for Scalar Invariance	*Step 8* Test for Factor Mean Invariance
	Step 5 Test for Error Variance Invariance	

Error of Approximation (RMSEA) (also see Browne and Cudeck, 1992), normed fit and non-normed fit indices (NFI and NNFI, respectively) (e.g. Bentler, 1990, Medsker, et al., 1994, Mulaik, et al., 1989), and Bentler's (1990) comparative fit index (CFI). Good overall fit indicates overall measurement invariance across cultural groups.

If poor fit is indicated, further tests concerning a "series of increasingly restricted hypotheses in order to identify the source of nonequivalence" (Byrne, 1989: 126) need to be conducted. Such additional tests are detailed in the following sections.

2) *Test for Configural Invariance:* Examination of configural equivalence is based on Thurstone's (1947) principle of factorial structure invariance and tests whether the same pattern of zero and nonzero loadings, but not their particular magnitude, are invariant across cultural groups (Steenkamp and Baumgartner, 1998). In other words, the test for configural invariance assesses whether different cultural groups have a similar pattern of factor loadings. Here, researchers need to impose equality constraints on the pattern of factor loadings across cultural groups, and test whether this pattern of free and fixed factor loadings is invariant across cultural groups (Vandenberg & Lance, 2000). Given that this test involves the analysis of separate measurement models (one

for each cultural group), researchers need to use SEM model comparison procedures, such as the chi-square difference test.

For example, a researcher comparing two cultural groups would need to run two measurement models: one freely estimated and the other with equality constraints imposed based on the factor structure of the first measurement model. Briefly, the procedure to test an invariant pattern of factor loadings based on the chi-square model comparison procedure is conducted as follows: (1) calculate the chi-square value and degrees of freedom for both models using a structural equation modeling software application, such as LISREL (Jöreskog and Sörbom, 2001); (2) calculate the difference between the chi-square values and degrees of freedom between the two models; (3) calculate the p-value for the chi-square difference between the two models based on the difference in the degrees of freedom between both models; and (4) compare the calculated p-value to a pre-defined alpha-level. This analysis leads to one of two scenarios:

> *Scenario 1* A significant chi-square difference implies that factor loadings are not invariant across the cultural groups. In such a situation, the underlying constructs are different across groups. Hence, additional invariance tests or tests of group differences -- such as examination of latent group differences, or group differences in structural parameters-- are not meaningful given that the underlying constructs are themselves different across cultural groups (Vandenberg and Lance, 2000).
>
> *Scenario 2* A chi-square difference that is not significant implies factor loadings are invariant across the cultural groups. Hence, respondents from each cultural group employ the same conceptual frame and therefore can be compared meaningfully (Vandenberg and Lance, 2000). Further invariance testing depends on the purpose of the study. If the purpose is to examine the basic meaning and structure of a construct or set of constructs across cultural groups, then testing for configural invariance is generally sufficient (Steenkamp and Baumgartner, 1998). However, if the purpose of a study is to compare the means of constructs across cultural groups, researchers need to employ additional invariance tests and also examine metric and scalar invariance for at least two items per construct (Steenkamp and Baumgartner, 1998).

3) *Test for Metric Invariance:* Metric invariance tests determine whether the magnitude of factor loadings is the same across cultural groups (Vandenberg and Lance, 2000). Metric invariance examines whether respondents from different cultural groups answered to the individual items in the same way, which provides a stronger test of measurement invariance than configural invariance (Steenkamp and Baumgartner, 1998). To assess metric invariance, equality constraints are imposed on the factor loadings to test whether the magnitude of factor loadings is equivalent across cultural groups. Separate measurement models (one per cultural group) and SEM model comparison methods are used to compare the groups. For example, a researcher comparing two cultural groups would need to run two measurement models: one that is freely estimated and a second that has imposed equality constraints based on the factor loadings of the first measurement model. The most frequently used method for such a model comparison is the chi-square difference test procedure (described in the discussion on configural invariance). Significance implies that the factor loadings are not invariant across cultural groups. No significance suggests metric invariance.

Cheung and Rensvold (1999) refined the test for metric invariance in the context of cross-cultural research and introduced the notion of item-level metric invariance. These authors contend that testing for basic (construct-level) metric invariance does not provide information on which specific items in a set of non-invariant items are nonequivalent (Cheung and Rensvold, 2002, 1999). One method that can be used to test for item-level invariance is the factor-ratio test (Cheung and Rensvold, 1999), which basically examines all pairs of items for metric invariance for each non-invariant construct (Cheung and Rensvold, 2002, 1999).

Hence, the factor-ratio test results in a number of invariant and a number of non-invariant items per non-invariant measurement scale. Researchers may either delete non-invariant items or remove the equality constraints from those items. The former method, however, is rather a-theoretical, particularly with pre-validated measures. Employing the latter method, non-invariant items are allowed to vary freely to test for partial (rather than full) metric invariance. Doing so can be meaningful as long as the number of items that is estimated freely is relatively small, and partial invariance of a focal measurement scale across cultural groups can be justified theoretically (Cheung and Rensvold, 2002, 1999).

4) *Test for Scalar Invariance*: Scalar invariance concerns "whether there is consistency between cross-national differences in latent means and cross-national differences in observed means [and] implies that cross-national differences in the means of the observed items are due to differences in the means of the underlying construct(s)" (Steenkamp and

Baumgartner, 1998: 80). This test is frequently used to test for invariance of systematic response/common method bias (Vandenberg and Lance, 2000). The null hypothesis stating that scalar invariance exists is tested and examined based on the chi-square model comparison procedure, which is discussed in the "Test for Configural Invariance" section.

An alternative method to the chi-square model comparison procedure is the examination of differences in other goodness of fit indices. Cheung and Rensvold (2002) note that although much research exists concerning goodness-of-fit indices in the context of single-model testing, fit indices --other than the likelihood ratio chi-square ($\chi 2$) statistic-- have not been considered in the context of multiple group comparisons. Although the literature has demonstrated that the chi-square test is dependent on sample size and number of parameters, and therefore should not be used in isolation in the context of single-model testing, researchers heavily rely on this approach when comparing multiple models. Cheung and Renvold identified three fit indices that are superior to chi-square due to their independence from sample size as well as their non-correlation with the goodness-of-fit statistics of the individual, overall models. These fit indices are Bentler's (1990) comparative fit index (CFI), McDonald's (1989) noncentrality index (NCI), and Steiger's (1989) gamma hat. A value of change smaller than or equal to -0.01, -0.001, and -0.02, respectively for the CFI, NCI, and gamma hat, indicates scalar invariance across cultural groups. This alternative model comparison procedure applies equally to the model comparisons of other invariance tests such as configural, metric, and error variance.

5) *Test for Error Variance Invariance:* This test examines the equivalence of unique variances and covariances of the residual variance (error terms) across cultural groups (Steenkamp and Baumgartner, 1998, Vandenberg and Lance, 2000) determining whether measurement reliability is equivalent across these groups. The error variance invariance test is conducted by imposing equality constraints on the items' uniquenesses across cultural groups, as well as --if testing for unique covariance invariance-- constraining uniqueness covariances to be equal across cultural groups (Vandenberg and Lance, 2000). The null hypothesis that error variance invariance exists is tested based on the logic mentioned for the preceding tests of configural, metric and scalar invariances. A significant model comparison indicates non-invariant error variances (and, if tested, non-invariant covariances) across cultural groups. An insignificant model comparison suggests equivalent error (and, if tested, invariant covariances) across cultural groups.

Cheung and Renvold (2002) categorized the aforementioned tests of configural, metric, scalar, and error variance invariance into the broad category of psychometric property invariance. These invariance tests constitute a prerequisite for the interpretation of additional invariance tests that concern another broad category, structural invariance, comprising tests of between-group differences in variances, covariances, and means. These additional invariance tests are generally conducted by researchers who have a more substantive interest in cross-cultural comparisons, given that these tests examine whether construct variability, construct relationships, and latent means are equivalent across cultural groups (Cheung and Rensvold, 2002, 1999, Vandenberg and Lance, 2000). In the following, these additional tests will be discussed briefly to provide the reader with a general idea about their purposes and underlying logic.

6) *Test for Construct Variance Invariance:* This test examines whether the construct variability is equivalent across cultural groups (Cheung and Rensvold, 1999, Vandenberg and Lance, 2000). In placing equality constraints between like factor variances, researchers can test for the null hypothesis of equal construct variances (Vandenberg and Lance, 2000). The same logic concerning model comparison and hypothesis testing, as mentioned for the other invariance tests, applies.

7) *Test for Construct Covariance Invariance:* In using this test, researchers are able to examine whether relationships among constructs are the same across cultural groups (Vandenberg and Lance, 2000). This test is typically used for an examination of the equality of factor intercorrelations, and so assesses whether specific relationships in a model hold across cultural groups (Cheung and Rensvold, 1999, Vandenberg and Lance, 2000). However, a debate exists concerning the rationale for this test, and some researchers believe that it does not add much value (e.g. Vandenberg and Lance, 2000). These researchers argue that if configural equivalence is established in earlier invariance tests then strength, pattern and construct/factor covariances are likely also invariant across cultural groups. Also, if the more stringent test of configural invariance indicates non-invariance of factor structures across cultural groups, then the interpretation of a significant test of construct covariance invariance as support for equivalent construct relationships across cultural groups is not very meaningful (Vandenberg and Lance, 2000).

8) *Test for Factor Mean Invariance:* Lastly, the test of factor mean invariance assesses the equality of factor means across cultural groups. Similar to the logic of analysis of variance, this test involves an initial overall test

of equivalence of factor means by testing a null hypothesis based on the chi-square difference model comparison approach which is discussed in the "Test for Configural Invariance" section. If a significant model comparison indicates non-invariance of factor means across cultural groups, then subsequent tests may be conducted to isolate specific differences between cultural groups. Doing so is similar to the logic of post-hoc tests in analysis of variance procedures.

By clearly considering the purpose of research, gaining a conceptual understanding about the underlying constructs and their meaning across cultures and carefully conducting all necessary invariance tests in the sequence described in this paper, researchers can overcome the invariance/equivalence pitfall. Neglecting such invariance tests or undertaking them in the wrong order may, however, have detrimental effects on the results of any analysis and their meaningful interpretation. Please also note that it is not meaningful to conduct invariance tests belonging to the third broad category, structural invariance tests (steps 6-8), prior to testing for invariance of the general psychometric properties of the measurement scale (steps 2-5).

PROBLEM III - THE SAMPLING PITFALL

Common pitfalls in sampling committed in domestic management research such as insufficient sample sizes, non-random sampling, and sampling of inappropriate subjects or even populations (e.g. Pedhazur & Pedhazur Schmelkin, 1991) equally apply to cross-cultural research, and some of these problems are exacerbated in cross-cultural settings

LEVEL OF ANALYSIS ISSUES

According to statistical analysis principles and psychometric theory, cross-cultural factor analysis methods require sample sizes of at least two observations per measurement item (Leung and Bond, 1989). However, a number of researchers advocate more stringent ratios, such as ten observations per measurement item (Nunnally, 1978). Operating on the ecological level of analysis using Hofstede's (1980) original measurement tool --the Values Survey Module (VSM) questionnaire, which assesses four cultural dimensions based on a total of 14 items-- one would need to collect data in at least 28 cultures if the minimum requirement of 2 observations per measurement item was used, or in as many as 140 cultures if Nunnally's recommendation was followed.. Whereas researchers primarily operating at the individual level of analysis may easily achieve such sample sizes, it is extremely difficult to achieve large sample sizes at the ecological level of analysis.

Multiple level research also generates the problem of achieving sample sizes according to the 'observation to item ratios' mentioned before for each level of analysis. Revisiting the aforementioned example, researchers operating at the ecological and individual levels of analysis using Hofstede's VSM (14 items) would need, at a minimum, 28 countries and 28 responses in each country (i.e. 784 responses in total). Applying Nunnally's more stringent ratio of 10 observations per measurement item would require data collection from 140 individuals in each of 140 cultures (which is to say, 19,600 responses in total).

COUNTRY SELECTION AND SAMPLING DESIGN

Another issue related to the sampling pitfall in cross-cultural management research is the misuse of national cultural dimensions, such as the ones developed by Hofstede. Although this problem has already been mentioned in the discussion on the levels of analysis pitfall, country selection and sample design are also facets of the problem. Hofstede has repeatedly emphasized that "[a]t times, my supporters worry me more than my critics" (2001: 73). This observation arises from issues concerning country selection and sampling design.

For example, researchers frequently investigate two countries representing two different levels on one cultural dimension, such as individualism. Researchers then sample a number of individuals or organizations in these two countries and consider the national cultural characteristic of individualism as an antecedent of their focal variable (for example, participatory management styles). However, it is difficult to conclude that both countries are sufficiently different on the individualism dimension to causally affect a difference in managerial styles as other dimensions (e.g. power distance, uncertainty avoidance, etc.) may also constitute important antecedents (Sivakumar and Nakata, 2001).

Accordingly, different management styles at the organizational level or different behaviors at the individual level of analysis cannot be predicted by ecological cultural dimensions with traditional analysis tools, but require tools capable of taking different levels of analysis into account, such as hierarchical linear modeling (HLM). Please note, however, that the use of HLM requires the need to collect data from a sufficient number of cultures in order to conduct appropriate statistical analyses.

HOW AND WHO TO SAMPLE

A last issue is how and who to sample in cross-cultural management research. The challenges of random sampling (time, cost, efforts and practical problems involved in achieving a random sample) are magnified in cross-cultural settings. Not only do researchers face larger spatial distances, but they also have to deal with different languages, dissimilar norms toward research and data collection, and other regulatory and cultural constraints.

Researchers should carefully consider how and who to sample in order to tap into the right population of respondents. Schwartz (1992), for example, sampled students and teachers in about 70 countries and justified this sampling design based on the notion that the educational system is likely to reflect both individual and society-level values. Inglehart's (1997) World Values Survey gathered data in several waves from the ordinary populations of the sampled countries (data were gathered by literally knocking on doors and interviewing ordinary people), given that this survey set out to uncover very general values such as happiness, traditional versus secular-rational values, and so on. GLOBE (House et al., 2004) surveyed about 17,000 middle managers in 951 organizations in 62 societies to learn about predominant leadership styles across countries. In summary, it is of utmost importance in cross-cultural management research to sample the right subjects from the right population(s) to avoid committing the sampling pitfall.

Strategies to Overcome the Sampling Pitfall

The first step in attempting to overcome sampling pitfalls involves an explicit recognition of the fact that the issue of an appropriate sampling design is paramount in international settings. Researchers need to understand the implications of levels of analysis (including multilevel research), country selection, and the question of how and who to sample for their cross-national sampling designs and data analyses.

Some researchers have advocated the use of nonparametric data analysis tools, which are not subject to such assumptions as the observation to item ratio that has been discussed before. For example, researchers may use Multidimensional Scaling (MDS) or Cluster Analysis in order to analyze cross-national data that do not meet the requirements of metric-type analysis techniques, such as cross-cultural factor analysis (Leung and Bond, 1989, Ronen and Shenkar, 1985). For example, a researcher using Hofstede's VSM (14 items) based on a sample size of 150 individuals from 15 cultures may use MDS to analyze the data. Briefly, MDS establishes a similarity matrix consisting of 14x14 correlations (due to the 14 items of the VSM). The major assumption of MDS is that the resulting 196 correlations provide a good estimate of the similarity between the 14 items.

Concerning the country selection issue, Sivakumar and Nakata (2001) developed a decision framework based upon which cross-cultural management researchers can select countries. Specifically, these authors developed algorithms that can be used to calculate the absolute and relative distance of two countries on Hofstede's culture dimensions; leading to a list of indexes for all culture dimensions and all possible pairs of cultures (e.g. a total of 1,378 pairs in case a researcher chooses two out of Hofstede's 53 sampled cultures). Using Sivakumar and Nakata's procedure, researchers may identify "optimal country pairs which maximize the difference on the focal variable while minimizing differences on the non-focal variables" (2001: 563).

In reference to the question of how and who to sample, researchers may use non-random sampling procedures, and should carefully identify the focal population and subjects within this population. For example, Brislin and Baumgardner (1971) were already concerned with the issue of sampling in international settings more than 35 years ago. These authors forward recommendations to improve non-random sampling in cross-cultural research, which include the use of a well-conceived sample description that should comprise "all the characteristics of the subjects and the environment which could potentially influence the results or their interpretation" (1971: 399). Furthermore, these authors note that researchers should indicate whether they categorized some individuals or variables (e.g. based on education level, socio-economic status), particularly in situations in which different cultural samples will be compared.

CONCLUSION

In summary, this paper has sketched out three frequently committed pitfalls in cross-cultural management research, which include level of analysis issues, testing for measurement invariance/equivalence across cultural groups, and sampling issues in international settings. Further, this paper has attempted to outline strategies to overcome these problems. Given that culture constitutes such an important and relevant construct in academic research, it is hoped that this paper furthers our understanding of the relevant issues that arise in empirical cross-cultural management research, and therefore contributes to the literature on international business.

REFERENCES

Bentler, P.M. (1990) 'Comparative fit indixes in structural models', *Psychological Bulletin*, **107**(2): 238-46.

Brislin, R.W. & Baumgardner, S.R. (1971) 'Non-random sampling of individuals in cross-cultural research', *Journal of Cross-Cultural Psychology*, **2**(4): 397-400.

Browne, M.W. & Cudeck, R. (1992) 'Alternative ways of assessing model fit', *Sociological Methods & Research*, **21**(2): 230-58.

Byrne, B. (1989) *A Primer to LISREL: Basic Applications and Programming for Confirmatory Factor Analytic Models*, Springer Verlag: New York.

Byrne, B.M. (1998) *Structural Equation Modeling: Basic Concepts, Application, and Programming*, Lawrence Earlbaum Associated, Inc.: Mahwah, NJ.

Castro, S.L. (2002) 'Data analytic methods for the analysis of multilevel questions: A comparison of intraclass correlation coefficients, rwg(j), hierarchical linear modeling, within- and between-analysis, and random group sampling ', *Leadership Quarterly*, **13**(1): 69-93.

Cheung, G.W. & Rensvold, R.B. (2002) 'Evaluating goodness-of-fit indexes for testing measurement invariance', *Structural Equation Modeling*, **9**(2): 233-55.

Cheung, G.W. & Rensvold, R.B. (1999) 'Testing factorial invariance across groups: A reconceptualization and proposed new method', *Journal of Management*, **25**(1): 1-27.

Dansereau, F., Yammarino, F., J. & Kohles, J., C. (1999) 'Multiple levels of analysis from a longitudinal perspective: Some implications for theory building', *Academy of Management Review*, **24**(2): 346-57.

Glick, W.H. (1985) 'Conceptualizing and measuring organizational and psychological climate: Pitfalls in multilevel research', *Academy of Management Review*, **10**(3): 601-16.

Hofmann, D.A. (1997) 'An overview of the logic and rationale of hierarchical linear models', *Journal of Management*, **23**(6): 723.

Hofstede, G. (1980) *Culture's Consequences: International Differences in Work-Related Values*, Sage: Beverly Hills, CA.

Hofstede, G. (2001) *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*, Sage: Thousand Oaks, CA.

Hofstede, G., Bond, M.H. & Luk, C.-l. (1993) 'Individual perceptions of organizational cultures: A methodological treatise on levels of analysis', *Organization Studies*, **14**(4): 483-503.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004) *Culture, Leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks: Sage Publications.

House, R., Rousseau, D.M. & Thomas-Hunt, M. (1995) 'The meso paradigm: A framework for the integration of micro and macro organizational behavior', in Cummings, L.L. and Staw, B.M. (eds.) *Research in Organizational Behavior*, JAI Press: Greenwich, CT.

Inglehart, R. (1997) *Modernization and Post-Modernization: Cultural, Economic, and Political Change in 43 Societies*, Princeton University Press: Princeton, NJ.

Jöreskog, K.G. (1971) 'Simultaneous factor analysis in several populations', *Psychometrika*, **36**: 409-26.

Jöreskog, K.G. & Sörbom, D. (2001) *LISREL 8: User's Reference Guide*, Scientific Software International: Lincolnwood, IL.

Klein, K.J., Tosi, H. & Cannella Jr., A.A. (1999) 'Multilevel theory building: Benefits, barriers, and new developments', *Academy of Management Review*, **24**(2): 243-48.

Klein, K.J., Dansereau, F. & Hall, R.J. (1994) 'Levels issues in theory development, data collection, and analysis', *Academy of Management Review*, **19**(2): 195-229.

Kline, R.B. (2005) *Principles and Practices of Structural Equation Modeling*, Guildford Press: New York.

Lenartowicz, T. & Roth, K. (2001) 'Does subculture within a country matter? A cross-cultural study of motivational domains and business performance in Brazil', *Journal of International Business Studies*, **32**: 305-25.

Lenartowicz, T. & Roth, K. (1999) 'A framework for culture assessment', *Journal of International Business Studies*, **30**(4): 781-98.

Leung, K., Bhagat, R.S., Buchan, N.R., Erez, M. & Gibson, C.B. (2005) 'Culture and international business: Recent advances and their implications for future research', *Journal of International Business Studies*, **36**: 357-78.

Leung, K. & Bond, M.H. (1989) 'On the empirical identification of dimensions for cross-cultural comparisons', *Journal of Cross-Cultural Psychology*, **20**(2): 133-51.

McDonald, R.P. (1989) 'An index of goodness-of-fit based on noncentrality', *Journal of Classification*, **6**: 97-103.

Medsker, G.J., Williams, L.J. & Holahan, P.J. (1994) 'A review of current practices for evaluating causal models in organizational behavior and human resources management research', *Journal of Management*, **20**(2): 439-64.

Mulaik, S.A., James, L.R., Van Alstine, J., Bennett, N., Lind, S. & Stilwell, C.D. (1989) 'Evaluation of goodness-of-fit indices for structural equation models', *Psychological Bulletin*, **105**(3): 430-45.

Nunnally, J.C. (1978) *Psychometric Theory*, McGraw Hill: New York.

Pedhazur, E.J. & Pedhazur Schmelkin, L. (1991) *Measurement, Design, and Analysis*, Lawrence Erlbaum: Hillsdale, NJ.

Peterson, M.F. (2005) 'Measurement metrics at aggregate levels of analysis: The SVS and its implications', Honolulu, HI: Presentation held at the 65th Annual Meeting of the Academy of Management.

Podasakoff, P.M., MacKenzie, S.B., Lee, J.-Y. & Podsakoff, N.P. (2003) 'Common method biases in behavioral research: A critical review of the literature and recommended remedies', *Journal of Applied Psychology*, **88**(5): 879.

Raykov, T. & Marcoulides, G.A. (2000) *A First Course in Structural Equation Modeling*, Lawrence Erlbaum: Mahwah, N.J.

Robinson, W.S. (1950) 'Ecological correlations and the behavior of individuals ', *American Sociological Review*, **15**: 351-57.

Ronen, S. & Shenkar, O. (1985) 'Clustering countries on attitudinal dimensions: A review and synthesis', *Academy of Management Review*, **10**: 435-54.

Rousseau, D.M. (1985) 'Issues of level in organizational research: Multi-level and cross-level perspectives', in Cummings, L. L. and Staw, B. M. (eds.) *Research in Organizational Behavior*, JAI Press: Greenwich, CT.

Schriesheim, C.A. (1995) 'Multivariate and moderated within- and between-entity analysis (WABA) using hierarchical linear multiple regression', *Leadership Quarterly*, **6**(1): 1-18.

Schwartz, S.H. (1992) 'Cultural dimensions of values - towards an understanding of national differences', *International Journal of Psychology*, **27**(3-4): 540

Sivakumar, K. & Nakata, C. (2001) 'The stampede toward Hofstede's framework: Avoiding the sample design pit in cross-cultural research ', *Journal of International Business Studies*, **32**(3): 555-74.

Steenkamp, J.-B.E.M. & Baumgartner, H. (1998) 'Assessing measurement invariance in cross-national consumer research', *Journal of Consumer Research*, **25**(1): 78-90.

Steiger, J.H. (1989) *EzPATH: Causal Modeling*, SYSTAT: Evanston, IL.

Steiger, J.H. & Lind, J. (1980) 'Statistically based tests for the number of common factors ', Paper presented at the annual meeting of the Psychometric Society, Iowa City.

Stening, B.W. & Everett, J.E. (1984) 'Attributions for promotion and demotion in the U.S. and India', *Journal of Social Psychology*, **122**: 151-56.

TheChineseCultureConnection. (1987) 'Chinese values and the search for culture-free dimensions of culture', *Journal of Cross-Cultural Psychology*, **18**: 143-64.

Thorndike, E.L. (1939) 'On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them ', *American Journal of Psychology*, **3**: 122-24.

Thurstone, L.L. (1947) *Multiple Factor Analysis*, University of Chicago Press: Chicago.

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): 4-69.