

MEASUREMENT ISSUES IN CROSS-NATIONAL RESEARCH

Jagdip Singh*

Case Western Reserve University

Abstract. Substantive inferences from cross-national studies have important implications for theory (e.g., because they reveal insights into generalizability and boundary conditions) and managerial practice (e.g., because they offer guidelines to MNC managers). However, few empirical studies attend to measurement issues involved in cross-national research, and still fewer recognize the risk of inferential errors that are likely to occur by overlooking measurement issues. We discuss four measurement issues, namely (1) standardized versus unstandardized coefficients, (2) the impact of measurement error and unequal reliability, (3) the overall error rate and simultaneous analysis and (4) construct equivalence. Using illustrative examples we demonstrate the nature of each of these problems, the likely impact they can have on substantive conclusions, and approaches for tackling these problems. Additionally, we reanalyze a recently published three-nation study by Dubinsky, Michaels, Kotabe, Lim and Moon [1992] to clarify these measurement concerns, highlight a methodological approach, and delineate the extent and severity of inferential errors. Our reanalysis shows that the interactive effects of these measurement issues are pervasive, complex and unpredictable. We close with implications for cross-national research in general.

We discuss four measurement issues in cross-national research and show that inattention to these issues is problematic and, in many instances, results in serious inferential and substantive errors. These issues include (1) the use of standardized versus unstandardized coefficients, (2) measurement error and unequal reliability, (3) overall error rate and (4) construct equivalence. While some of these issues have been discussed in the literature (*JIBS*, Fall 1983), our discussion contributes to the literature in three ways. First, none of the previous studies focus exclusively on measurement issues in cross-national research. Because these issues influence substantive inferences, a detailed,

*Jagdip Singh is Associate Professor, Division of Marketing, Weatherhead School of Management, Case Western Reserve University, Cleveland, Ohio. This research was partially accomplished when the author was visiting fellow at the Timbergen Institute, School of Economics, Erasmus University, Rotterdam, The Netherlands.

The author thanks three anonymous *JIBS* reviewers for their helpful comments throughout the review process. In addition, the author is indebted to Roy Howell for his generous comments and encouragement. Additional statistical details concerning the unrestricted and restricted models estimated and reported in this paper are available from the author.

Received: January 1994; Revised: August & December 1994; Accepted: December 1994

focused discussion of measurement problems is likely to be useful for researchers. Second, we discuss the measurement issues from an empirical and analytical standpoint. Problems as they occur in empirical research are highlighted, and specific analytical strategies are provided for tackling these problems, thereby providing practical guidelines for empirical research. Third, we illustrate the measurement issues and analytical strategies by using examples and reanalysis of a study previously published in *JIBS*. We utilize this study and other examples to highlight measurement issues in cross-national research, and for understanding their implications for substantive inferences. We begin with a discussion of the four measurement issues.

STANDARDIZED VERSUS UNSTANDARDIZED COEFFICIENTS

In understanding the relationship between two or more constructs (or variables), researchers may use either the standardized or unstandardized regression coefficients. These coefficients are rarely identical in most empirical studies and differ in important ways. Thus, this choice is not a matter of convenience. Rather, researchers must formulate an explicit, coherent and informed decision by evaluating the inherent characteristics of standardized and unstandardized coefficients.

As an illustrative example, consider two independent variables *compensation*, X_1 (say, measured in thousands of dollars), and *motivation*, X_2 (measured on a five-point Likert scale), and one dependent variable, *job satisfaction*, Y (also measured on a five-point Likert scale). Assume that data are collected from two samples of salespeople with sizes N_1 and N_2 in a cross-national context (say, U.S. and Korea). Table 1 displays the basic statistics from these hypothetical data. The relationship among these variables is given by the regression equations:

$$Y_{11} = \beta_{01} + \beta_{11} X_{11} + \beta_{21} X_{21}; \text{ for sample 1 (U.S.) and} \quad (1)$$

$$Y_{22} = \beta_{02} + \beta_{12} X_{12} + \beta_{22} X_{22}; \text{ for sample 2 (Korea).} \quad (2)$$

Here β_{01} and β_{02} are the intercept terms, while β_{11} , β_{21} , β_{12} and β_{22} are the unstandardized coefficients. Assume that regression parameters are $\beta_{11} = .08$, $\beta_{21} = .30$, $\beta_{12} = .04$, and $\beta_{22} = .30$. However, the standardized coefficients, $s\beta$ s can be derived computationally from unstandardized coefficients by utilizing the standard deviations of relevant X and Y variables.¹ Based on the statistics provided in Table 1, it is apparent that $s\beta_{11} = .80$ (i.e., $.08*(10/1)$). Likewise, we can compute $s\beta_{21} = .30$, $s\beta_{12} = .20$ and $s\beta_{22} = .60$.

Some researchers prefer standardized coefficients on the grounds of their (1) interpretability, (2) common metric or "scale" and/or (3) "emic" comparison standard. First, the interpretability argument draws on the notion that, under some conditions, the square of the standardized coefficient is the variance explained in the dependent variable.² For instance, $(s\beta_{11})^2$ (i.e., $.64$) equals the

amount of variance in job satisfaction that is explained by compensation in sample 1. The unstandardized coefficients do not yield such a direct assessment.

TABLE 1
Hypothetical Data to Illustrate Measurement Issues in
Cross-National Research

Variable	← Sample 1 (US) →			← Sample 2 (Korea) →		
	Mean	Standard Deviation	Reliability ^a	Mean	Standard Deviation	Reliability ^a
Compensation (X_1) (in '000 \$)	50	10	1.0	20	5	1.0
Motivation (X_2) ^b (1 = low; 5 = high)	4.0	1.0	.70	3.0	2.0	.90
Job Satisfaction (Y) ^b (1 = low; 5 = high)	3.5	1.0	.80	3.0	1.0	.80

^a This is an estimate for the Cronbach's *alpha* reliability coefficient.

^b This is assumed to be a perceptual construct that is measured with multi-item scale.

Second, the common metric or “scale” notion argues that the standardized coefficients bring the disparate regression estimates to a common metric, irrespective of the scale utilized to measure the independent variables. Thus, $s\beta_{11}$ (.80) and $s\beta_{21}$ (.30) are on a common metric regardless of the measurement scale utilized for X_1 and X_2 . This common metric allows a comparative analysis of the effects of compensation and motivation such that it is possible to infer that for sample 1 compensation has a stronger effect than motivation on job satisfaction of salespeople (i.e., $s\beta_{11} > s\beta_{21}$). In contrast, unstandardized coefficients do not always reveal such comparative insights.

Third, the standardized coefficients reflect an “emic” comparison standard since the individual regression coefficients are adjusted on the basis of *within*-sample variability. As such, $s\beta_{11}$ is obtained by “adjusting” β_{11} by a factor that represents the variability of X_1 and Y within sample 1 (i.e., s_{x11}/s_{y11}), while the “adjustment” factor for $s\beta_{12}$ is based on the variability within sample 2 (i.e., s_{x12}/s_{y22}). This within-sample adjustment ensures that standardized coefficients have the same metric within a sample but *not* across samples.

On the other hand, the unstandardized coefficient is thought to be desirable due to (1) comparability across samples, (2) structural invariance, and (3) “etic” comparison standard. First, the notion of across-sample comparability implies that a valid comparative analysis of corresponding regression coefficients can be conducted only by utilizing unstandardized coefficients. As such, β_{21} and β_{22} can be utilized to draw valid comparative inferences, but

not $s\beta_{21}$ and $s\beta_{22}$. This is because by adjusting each coefficient by its own within-group variability, standardized estimates in effect eliminate any across-group differences on account of disparate variances. Thus, it is inappropriate to conclude that the effect of motivation on job satisfaction is twofold stronger in the Korean sample than in the U.S. sample (i.e., $s\beta_{22} = 2*s\beta_{21}$). Rather, the valid inference is that the effect of motivation is identical in both samples (i.e., $\beta_{21} = \beta_{22} = .30$).

Second, by structural invariance we imply that unstandardized coefficients represent structural parameters that are likely to remain invariant (statistically) for estimates obtained from different samples. In other words, if one were to draw several different salesperson samples say from the U.S., the unstandardized regression coefficient estimated for the link $X_1 \longrightarrow Y$ from each such sample will statistically equal β_{11} .³ By contrast, the standardized coefficients will vary from sample to sample, and may differ remarkably from $s\beta_{11}$ in some, if not most, samples.

Third, and finally, because they are unadjusted for within sample variability, the unstandardized coefficients reflect an "etic" comparison standard. Thus, the use of unstandardized coefficients presumes that the constructs and/or variables achieve "equivalence" for cross-national samples and, consequently, their scales of measurement are directly comparable. As such, in the hypothetical example we presume that, without any loss of validity, the compensation of Korean salespeople can be computed in U.S. dollars so that a common, equivalent scale can be utilized to measure compensation in the U.S. and Korean samples. If this is not feasible, construct equivalence is not achieved for the compensation variable and across-nation comparisons are not permissible.

Overall, it is apparent that, if the researcher's aim is to compare regression coefficients *across* two or more groups and the variables satisfy the etic criterion of construct "equivalence" (to be discussed), unstandardized coefficients should be utilized for interpretation and substantive inference. Conversely, if the variables lack "equivalence" and/or the objective is to only draw *within*-group comparisons, one should utilize standardized coefficients. We are not alone in our conclusion about standardized coefficients. Bollen [1989, p. 126] warns about the "hazard of comparing standardized coefficients . . . for the same variable across different groups" and recommends that "in general, comparisons of a variable's influence in different groups should be made with unstandardized coefficients." Likewise, Alwin (1988, p. 20) observes that "when the objective is *to compare the magnitudes of coefficients for a given variable in equations specified in different populations, the general practice is to compare the regression coefficients in their original metric*, rather than to rely on standardized units" [emphasis in original].

MEASUREMENT ERROR AND UNEQUAL RELIABILITY

Davis, Douglas and Silk [1981] have called measurement error or unreliability a “hidden threat” to cross-national research because of three interrelated propositions. First, in most empirical research, despite the best efforts of researchers, some error occurs in the process of measuring constructs. This error is termed measurement error and is indexed by a reliability coefficient that varies from 0 (100% measurement error) to 1 (0% measurement error). Second, when the *same* construct is measured in different cross-national contexts, reliability estimates are rarely equivalent. Instead, these estimates are likely to vary, often significantly, for cross-national data. For instance, Parameswaran and Yaprak [1987] empirically demonstrate reliability scores varying from .54 to .85, and which are sensitive to the nature of the construct measured, nationality of the respondents and country-of-origin effects. Third, and most critical, substantive relationships among constructs must be “adjusted” for unequal reliability in order to draw valid inferences. Thus, Davis et al. [1981, p. 99] suggest that, “what might appear to be a cross-national difference could turn out to be solely a reflection of variations in the reliability of the underlying measurements.” The “adjustment factor” is based on appropriate attenuation formulas (see Nunnally [1978, pp. 219–20]). For correlational data and Cronbach’s *alpha* (α) reliability estimates, the attenuation formula is as follows:

$$\rho_{xy} = r_{xy}/(\sqrt{r_{xx}} \cdot \sqrt{r_{yy}}), \quad (3)$$

where r_{xx} and r_{yy} are the reliability estimates, r_{xy} is the observed correlation, and ρ_{xy} is the corrected correlation for X and Y respectively. Thus, the “adjustment factor” is $1/(\sqrt{r_{xx}} \cdot \sqrt{r_{yy}})$.

The hypothetical data in Table 1 helps illustrate this adjustment factor. The “adjustment factor” for the relationship between X_1 and Y is 1.12 for both samples. In contrast, the “adjustment factor” for the relationship between X_2 and Y is 1.34 and 1.18 for the U.S. and Korean samples respectively. Thus, if the observed correlations between motivation (X_2) and job satisfaction (Y) were .45 and .51 in the U.S. and Korean samples respectively, it would be incorrect to conclude that a cross-national difference exists. Rather, the valid inference is that the relationship is equivalent across the two samples and the corrected correlation between X_2 and Y is .60 (i.e., $1.34 \cdot .46 = 1.18 \cdot .51 = .60$). Likewise, if the relationship between compensation (X_1) and job satisfaction (Y) was .55 in both samples, it would still be incorrect to surmise that the focal relationship is equivalent cross-nationally and equals .55. Instead, the valid inference is that, while the relationship is equivalent, the strength of the relationship equals .62 in both samples (i.e., $1.12 \cdot .55 = .62$).

Three points concerning the adjustment factor are noteworthy. First, while we present examples with correlational data, the notion of an adjustment factor

on account of unequal reliability applies with equal force for standardized and unstandardized coefficients. However, the corresponding adjustment factors differ somewhat from equation (3). Second, in a “bivariate” case (i.e., involving just two variables), the adjustment factor is always greater than 1 as long as reliability is less than 1.00. As such, *observed* bivariate correlations and regression coefficients are likely to underestimate the strength of the relationship between two constructs. Third, in a “multivariate” case (i.e., involving multiple independent variables and one or more dependent variables), the impact of measurement error is more pervasive and intricate than in the bivariate case. Although the mathematical formulas and statistics are somewhat involved (see Bollen [1989]), the key principles can be illustrated with the hypothetical example in Table 1. Thus, in the multivariate instance of estimating the joint effects of compensation (X_1) and motivation (X_2) on job satisfaction (Y), the impact of measurement error is such that the adjustment factor, say for the $X_2 \rightarrow Y$ relationship, will (1) be based on the reliability coefficients of X_2 , Y and X_1 in each sample, and (2) not necessarily be greater than 1.00. Thus, the corrected regression coefficients may be greater, smaller or equal to the corresponding observed coefficients depending on the adjustment factor.

OVERALL ERROR RATE

It is usually not sufficient to establish that some significant cross-national differences exists. Rather, the substantive interest lies in delineating the source (i.e., which pairs of nations contribute to this difference) and direction (i.e., which nation has the higher mean/coefficient) of the significant differences. Thus, for instance, in a recent three-nation comparative study of industrial salespeople, Dubinsky, Michaels, Kotabe, Lim and Moon [1992] (hereafter DMKLM) delineate the source and direction of differences in estimated regression coefficients by performing twenty-seven pairwise tests (their Table 4). The twenty-seven pairwise tests arise due to nine relationships posited for the effects of role stressors on work outcomes and three sets of tests considering each pair of countries involved (i.e., U.S./Japan, U.S./Korea and Japan/Korea). Such tests are usually referred to as “multiple comparison” tests.

Three principles dominate discussions of multiple comparisons. First, it is inappropriate to perform statistical tests for each pairwise comparisons by utilizing conventional levels of significance (say, 95%). This is because multiple comparisons inflate the overall Type-I error rate – the probability of rejecting the null hypothesis when in fact the null is true. For instance, in the case of twenty-seven pairwise comparisons each at 95% confidence level, the overall error rate is 75%.⁴ This equals a confidence level of 25% and considerably increases the chances of finding spurious cross-national differences.

Second, depending upon the nature of comparisons (e.g., post-hoc or a priori), various procedures are available to control overall error rate [Klockars and Sax 1986]. A common procedure for a priori comparisons is to set the confidence level for each pairwise test to $[1 - \alpha/n]$ where α is the desired overall error rate and n is the number of multiple comparisons. In DMKLM's study if the overall error rate is to be 5%, each of the twenty-seven pairwise tests must be evaluated at a confidence level of 99.82% (i.e., $p = .0018$).

Third, it is almost always more appropriate to utilize a multivariate test (e.g., omnibus F -test) to control overall error rate than to adjust the confidence level. A multivariate test examines all pairwise comparisons simultaneously. This approach is superior because it utilizes more information (e.g., correlations among the variables) and controls overall error rate within predetermined bounds. Indeed, if and *only* if the multivariate test indicates that at least one cross-national pair is significantly different, univariate tests can be performed to locate the source and direction of significant differences [Hummel and Sligo 1971]. Thus, for twenty-seven pairwise tests, valid cross-national inferences demand (1) a multivariate test that simultaneously examines all twenty-seven comparisons, and (2) follow up univariate tests only if the multivariate test is significant.

CONSTRUCT EQUIVALENCE

The notion of "construct equivalence" is rooted in the etic perspective and involves three aspects. First, it examines if a given construct serves the same function and is expressed similarly (i.e., in terms of attitude or behaviors) in different cross-national contexts. This is often referred to as functional and conceptual equivalence. Second, it explores if the construct or scale items, response categories and other questionnaire stimuli (e.g., instructions) are interpreted similarly in cross-national settings. This is usually termed instrument equivalence. Finally, it examines if each scale item measures the underlying construct equivalently in cross-national data. This occurs after the data are collected and is referred to as measurement equivalence. Considerable literature exists concerning the problems that stem from a lack of, and methodological approaches that help achieve, construct equivalence [Green and White 1976; Brislin 1976; Sekaran 1983]. Space limitations permit only a brief review of selected issues that are often neglected in cross-national research.

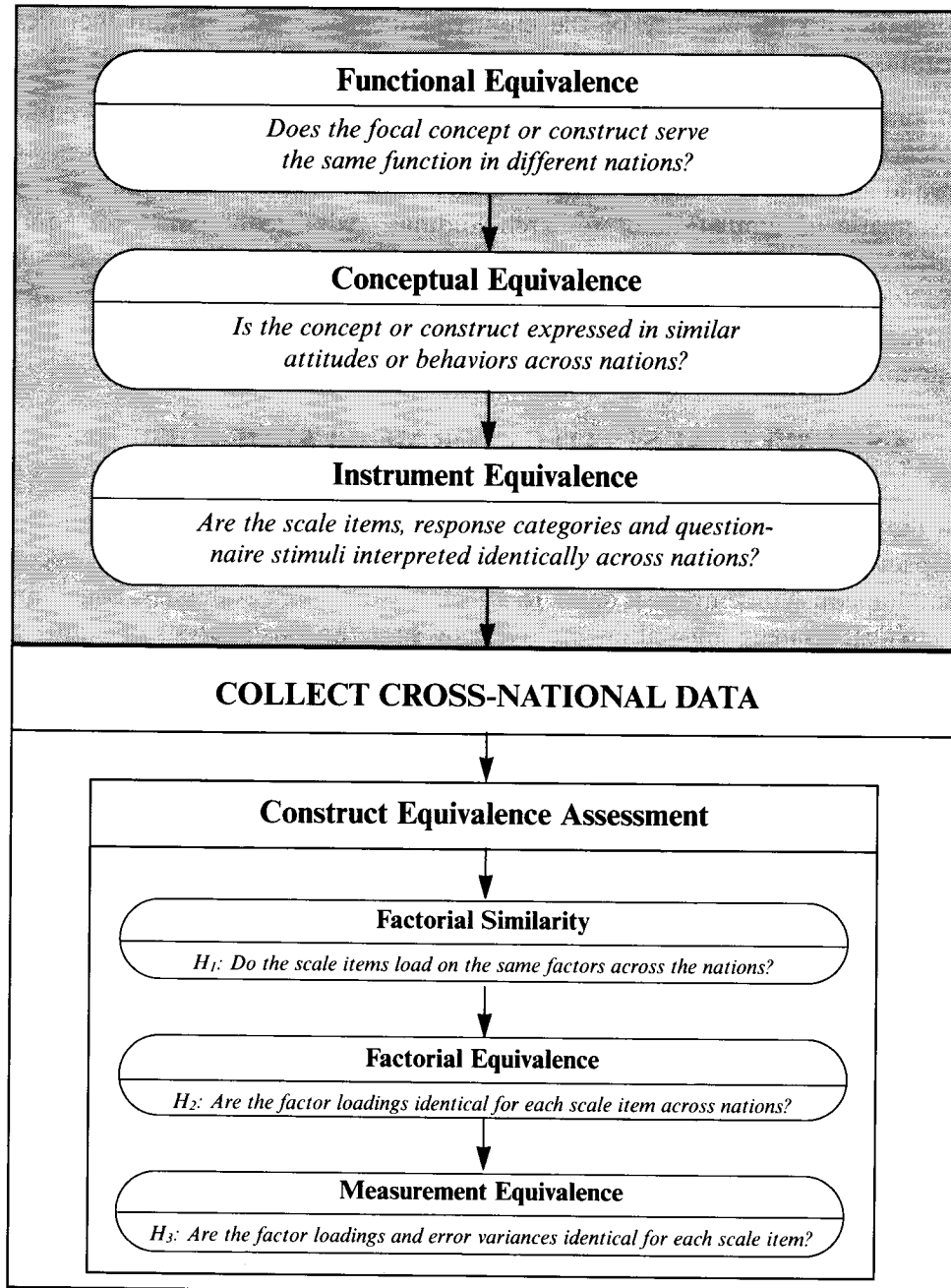
Most researchers agree that a lack of construct equivalence threatens the validity of substantive inferences in cross-national research [Adler 1983]. The hypothetical example of Table 1 serves to illustrate these threats. Suppose the motivation (X_2) construct lacks equivalence across the U.S. and Korean samples. In particular, assume that for a subset of Korean respondents, the motivation scale items appear biased (i.e., lack of instrument equivalence)

such that these respondents tend to respond less positively than they would have done otherwise (i.e., resulting in lack of measurement equivalence). In this case, it is inappropriate to substantively examine the mean values for X_2 in Table 1 and conclude that Korean salespeople are less motivated than U.S. salespeople. Rather, the valid inference is that the mean value for the Korean sample is underestimated and a lack of construct equivalence has confounded our ability to draw substantive conclusions. Likewise, the regression coefficient for the effect of X_2 on Y will be biased because the motivation construct is not equivalent across the two samples. Thus, valid substantive inferences are predicated on satisfactory attainment of construct equivalence in a cross-national study.

It is especially unfortunate if an impression persists among cross-national researchers that the degree of construct equivalence *cannot* be assessed in empirical research. This is because early work in assessing construct equivalence dates back to 1970s (cf. Joreskog [1971]) and several excellent "how-to" expositions are available (e.g., see Drasgow and Kanfer [1985]; Bollen 1989]). In addition, useful tutorials of different methodologies for evaluating measurement equivalence have begun to appear in *JIBS* (see Mullen [1995], this issue). While we refer readers to these sources for a detailed discussion, three principles need mention. First, procedures for assessing construct equivalence are implemented *after* the cross-national data are collected. Figure 1 reflects this notion by depicting various issues involved in construct equivalence. Specifically, Figure 1 shows that issues of functional, conceptual and instrument equivalence need to be addressed *before* cross-national data collection, while equivalence assessment is only possible *after* the data collection stage. This should not be taken to imply that procedures that are utilized *before* data collection (e.g., functional/conceptual/instrument equivalence) are of less importance. Rather, the aim of *after* procedures is to probe the degree of success of *before* procedures.

Second, assessment procedures provide statistical information by utilizing simultaneous multi-group factor analysis to test several hierarchical hypotheses that correspond to increasing degrees of construct equivalence. In Figure 1, we show three such hypotheses. The first hypothesis examines factorial similarity across nations. By *factorial similarity* we imply that scale items load on the same factor (or construct) in cross-national data. This condition is necessary but not sufficient for construct equivalence. The second hypothesis – *factorial equivalence* – tests if each scale item has the same loading (within statistical bounds) and on the same factor in cross-national data. Many researchers incorrectly regard this as a sufficient condition for construct equivalence. The third hypothesis represents the highest degree of construct equivalence and posits that factor loadings and error variances are identical for each scale item. This is the *measurement equivalence* hypothesis.

FIGURE 1
Construct Equivalence Issues in Cross-National Research



Third, the notion of construct equivalence assessment is *not* an all-or-none concern. Rather, these procedures can be utilized to identify offending scale item(s) that violate construct equivalence expectations. This power can be utilized in two ways: (1) to eliminate offending items in the focal study so that

valid substantive conclusions can be drawn, and (2) to target items that need further development in future cross-national research. Moreover, construct equivalence is *not* a bounded concern. Rather, it represents a continual concern with the validity of the empirical measures utilized and includes notions of discriminant, predictive and nomological validity assessment.

In sum, cross-national research requires clear evidence of construct equivalence. Failure to assess construct equivalence increases the chances of invalid substantive inferences, perpetuates unsound measures and hinders the systematic accumulation of research findings. Admittedly, construct equivalence is a complex topic requiring familiarity with rigorous research methodologies. However, as noted by Adler [1983] and Sekaran [1983] almost a decade ago, the perils of ignoring construct equivalence are formidable and warrant a serious and thorough analysis of such issues.

AN ANALYTICAL STRATEGY FOR MEASUREMENT ISSUES: MODEL EQUIVALENCE

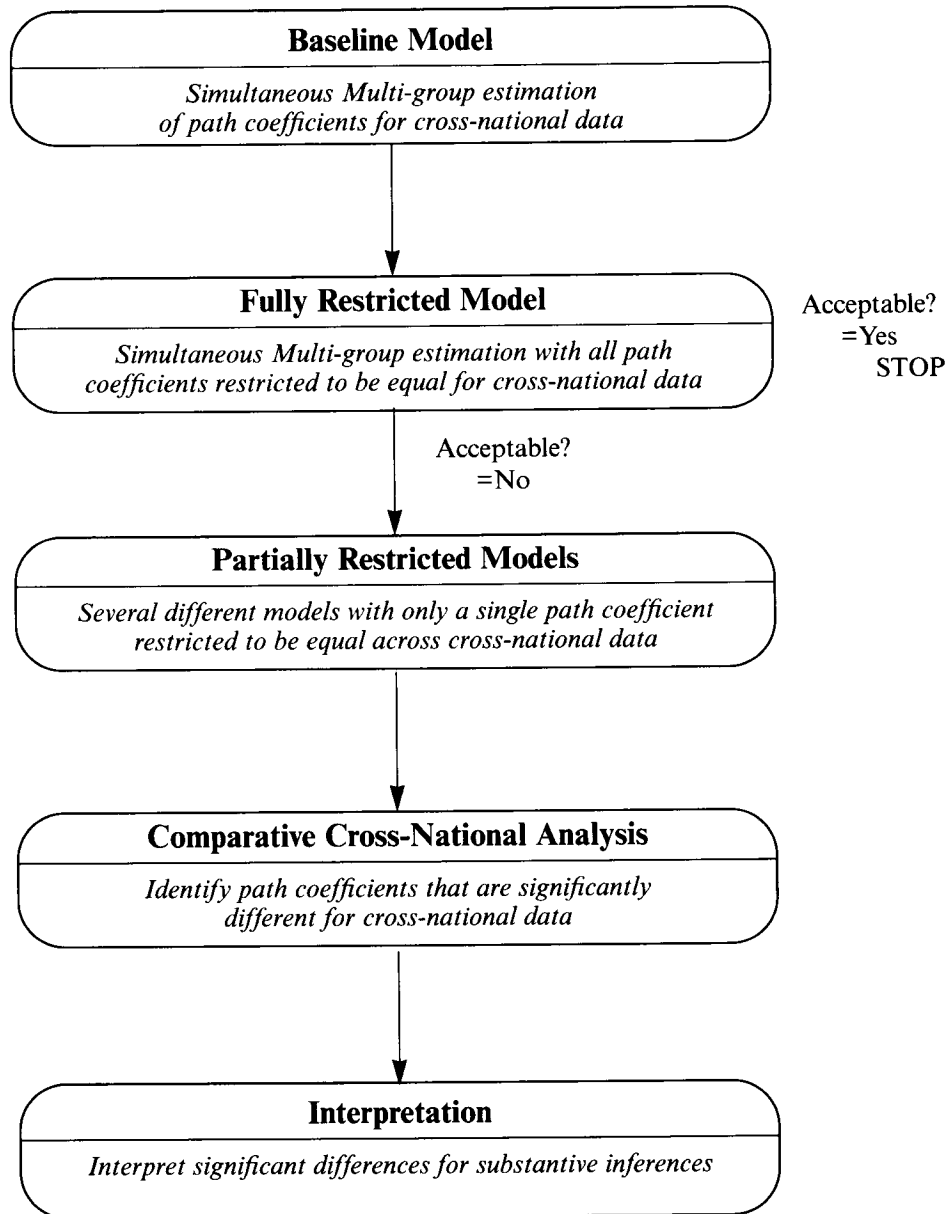
In the preceding sections, the four measurement issues and their substantive impact are discussed separately. In reality, however, these issues often occur simultaneously in empirical data. To effectively tackle measurement issues, an analytical strategy must allow multivariate, simultaneous analysis of multiple cross-national data sets. Thus, in the hypothetical example of Table 1, it would be necessary to develop an analysis strategy that simultaneously (1) estimates the regression coefficients for both cross-national data sets, (2) adjusts for unequal reliability for all constructs involved, (3) allows multivariate and bivariate comparisons with some control on overall error rate, and (4) offers possibilities for testing construct equivalence (if necessary). In essence, this strategy boils down to testing for *model equivalence* across cross-national data sets. Specifically, for issues (1), (2), and (3), a path model described by simultaneous regression equations (1) and (2) must be tested for equivalence in U.S. and Korean samples. In regard to issue (4), however, a factor model including the hypothesized factor loadings is tested for equivalence for the cross-national samples. In advanced structural models, path and factor models may be tested simultaneously for equivalence [Drasgow and Kanfer 1985; Bollen 1989]. Such models utilize the multi-trait multi-method (MTMM) nature of cross-national data (where "traits"=constructs, "methods"=nations) to employ a comprehensive apparatus for testing measurement issues [Byrne 1989; Cole and Maxwell 1985].

One promising analytical strategy is based on the method of multi-group Latent Variable Structural Equation (LVSE) modelling via a software such as LISREL or EQS [Bollen 1989]. Compared to regression procedures, the use of LVSE-based approach has several advantages. First, it allows for a simultaneous examination of a system of hypothesized equations involving

multiple dependent variables. As such, the interrelationships among hypothesized antecedents of, say, job satisfaction and performance can be examined simultaneously. Second, as a multi-group approach, it allows for a simultaneous estimation of a system of equations in multiple datasets, such as the U.S. and Korean data. Third, it provides multivariate goodness-of-fit statistics including an overall *chi-square* statistic, and several goodness-of-fit indices such as normed (NFI), nonnormed (NNFI) and comparative-fit indices (CFI). A *nonsignificant* χ^2 and fit indices greater than .95 are indicative of a model that fits the data reasonably well [Bollen 1989]. Fourth, it is feasible to set the measurement error for each construct in each cross-national data set by using its estimated *alpha* reliability via an “adjustment” factor [Bollen 1989].⁵ This procedure is desirable because the estimated path coefficients are corrected for unequal reliability. Finally, this approach allows for “restricted” models with systematic constraints on hypothesized relationships across the cross-national datasets. A key implication is that models can be tested that restrict all or selected path coefficients to be equal for cross-national datasets. This is useful for comparative analysis and yields a reasonable control on overall error rate.

Specifically, Figure 2 depicts a five-step analytical strategy based on the LVSE approach that is suited to cross-national research. For sake of clarity, we describe testing for path model equivalence. Factor model equivalence for testing construct equivalence follows a similar strategy but substitutes factor loadings for path coefficients. First, an “unrestricted” model is estimated in which path coefficients are allowed to vary across the cross-national datasets. In the second step, a “fully restricted” model is estimated by restricting each path coefficient to be equal for cross-national data sets. As such, the “fully restricted” model is based on the notion of *invariance* of model relationships in cross-national settings. A comparison of the multivariate goodness-of-fit statistics for the “unrestricted” and “fully restricted” models – based on a *chi-square* difference test⁶ – yields evidence for the plausibility of the invariance hypothesis. Because the *chi-square* difference tests for all path coefficients simultaneously, its multivariate nature offers control on overall error rate. Specifically, if this test is *nonsignificant*, it suggests that the “fully restricted” model is acceptable and *none* of the coefficients differ for the cross-national datasets. Conversely, if the test is significant, it indicates rejection of the invariance hypothesis and the presence of significant cross-national differences. In this case, the third step can be implemented involving several “partially restricted” models that restrict path coefficients one-at-a-time to be equal for cross-national data. By repeated use of the *chi-square* difference test, it is possible to conduct comparative analysis and delineate coefficients that differ significantly. In the final step, these significant path coefficients can be interpreted in the context of substantive theory to yield inferences for theoretical work and managerial practice.

FIGURE 2
A LVSE-Based Analytical Strategy for Tackling Measurement Issues in Cross-National Research: Testing for Model Equivalence



ILLUSTRATIVE EXAMPLE: A REANALYSIS OF DMKLM's DATA

We illustrate the measurement issues by a reanalysis of DMKLM's [1992] study of the effects of role stressors on work outcomes for comparable samples of salespeople from United States (hereafter "U.S."), Japan and South Korea (hereafter "Korea"). For several reasons, DMKLM's study is

suited to an illustrative reanalysis. First, this study involves (1) multiple constructs measured by (2) multi-item measures with unequal reliability, and (3) multiple comparisons based on a priori hypotheses for cross-national differences. As such, DMKLM's research confronts several of the measurement issues discussed herein and, consequently, offers an appropriate context to illustrate key ideas. Second, cross-national comparisons and procedures used in DMKLM's study warrant further discussion and analysis. A reanalysis of their data suggests different conclusions. We identify key distinctions between DMKLM's findings and the reanalysis, and draw its implications for sales management research in particular, and cross-national studies in general.

The Original DMKLM Analysis

DMKLM's research posits interrelationships among two *independent* constructs, Role Conflict (RC) and Role Ambiguity (RA), and three *dependent* constructs, Job Performance (JP), Job Satisfaction (JS) and Organizational Commitment (OC) captured by nine hypothesized paths based on the theoretical framework of Kahn, Wolfe, Quinn, Snoek and Rosenthal [1964], and substantiated by empirical research conducted mainly in the U.S. DMKLM utilize salespeople survey data with response rates ranging from 34.7% (in Korea) to 64.1% (in U.S.). Most measures utilized represent standardized scales that have been used extensively in the U.S. Thus, using the English version as the common anchor, DMKLM develop "equivalent" Japanese and Korean versions through a process of translation and back-translation. The estimated reliability coefficients vary, sometimes considerably, across the three data sets. For instance, reliability coefficient for role ambiguity varies from .62 for Japanese data to .80 for U.S. data. DMKLM do not use this variability in subsequent analysis.

DMKLM estimate the posited relationships by using several regression analyses. No account is taken of measurement error. Although it is not clear from their study, DMKLM appear to adopt an *etic* perspective in hypotheses testing and drawing implications from their research. This is because DMKLM are interested in cross-national comparisons, as they observe that the estimated coefficients are likely to differ in "*degree*, if not in *kind*" (p. 83) in the U.S., Japan and Korean samples. In accord with this, they test the standardized coefficients for each bivariate "pairs" of countries to infer if a particular relationship varies significantly across the three samples. DMKLM conclude that (1) none of regression coefficients for U.S. and Korean samples differ significantly, and (2) two of the regression coefficients differ significantly for the U.S./Japanese and Japanese/Korean samples. Specifically, for the U.S./Japanese comparison, DMKLM report that the effect of role conflict on role ambiguity is positive (as expected) and significantly larger for the U.S. sample (.63) than the Japanese sample (.36) and the role conflict-job satisfaction link is negative (as expected) and about twofold larger for the U.S.

sample ($-.58$) than the Japanese sample ($-.29$). For the Japanese/Korean comparison, DMKLM concluded that the effect of role conflict on job satisfaction is negative (as expected) and higher for the Korean sample ($-.50$) than the Japanese sample ($-.29$). In contrast, the performance-satisfaction link is positive (as expected) but several times larger for the Japanese sample ($.33$) than the Korean sample ($.06$).

Of equal significance are the hypotheses that were not supported. DMKLM had posited that role conflict would have a significant, negative effect on performance, and that this effect would be weaker in Japan than in the U.S. or Korea. DMKLM found the effect of role conflict on performance to be positive in all cases, and significant only for Japanese data. In regard to the role conflict-commitment link, DMKLM had posited that role conflict would have a significant, negative effect on commitment in each sample. However, DMKLM found that this effect was *not* significant for the Korean sample, but achieved significance in the other two samples.

The LVSE-Based Reanalysis

The analytical strategy of Figure 2 was utilized for reanalysis of DMKLM data. Covariance matrices from three samples were computed and input into EQS for estimating the various models.⁷ Maximum likelihood (ML) estimates of path coefficients were obtained via EQS. The measurement error of each construct was fixed using the adjustment formula and reliability estimates from DMKLM (Note 6). ML estimates are desirable as they are unbiased and yield minimum variance. However, certain assumptions of the ML method, including normality, may present limitations.

Tests of Model Equivalence. The *unrestricted* model yields the following multivariate goodness-of-fit statistics: $\chi^2 = 14.07$, $df = 6$, $p = .03$, NFI = .99, NNFI = .96, and CFI = .99.⁸ The fully restricted model has comparable statistics as follows: $\chi^2 = 56.47$, $df = 24$, $p < .001$, NFI = .94, NNFI = .95, and CFI = .96. The χ^2 difference statistic is 42.40 (i.e., $56.47 - 14.07$) with 18 degrees of freedom (i.e., $24 - 6$). This difference statistic is highly significant (i.e., $p < .001$) indicating that the fully restricted model is unacceptable and rejects the invariance hypothesis. Thus, at least one path coefficient differs significantly for the cross-national data. This accords with DMKLM's results.

Thereafter, several partially restricted models were estimated to test if the hypothesized path coefficients were invariant across the three cross-national datasets (see Table 2). For each hypothesized path (e.g., RC \longrightarrow RA), two models were compared: (1) a baseline, unrestricted model in which the path coefficient was allowed to vary, and (2) a restricted model in which the path coefficient was fixed to be equal across the three datasets (i.e., invariant). A χ^2 difference statistic was computed to test the invariance hypothesis and obtained by taking the difference between the model χ^2 for the unrestricted

and restricted models. This χ^2 difference statistic along with its significance level is shown under the “reanalysis” column. A *nonsignificant* result at $p > .05$ indicates that the focal path coefficient is *not* significantly different across the three datasets. The left-most column presents the specific path coefficient that was restricted to be equal across the three datasets. Analogous results from DMKLM are also included.

Overall Differences

Table 2 reveals critical inferential differences between DMKLM’s results and the reanalysis. Specifically, it appears that DMKLM inadvertently commit at least one Type-I error (i.e., concluding that there is a significant difference when in fact there was not) and one Type-II error (i.e., concluding that there was no difference when in actuality there was a significant difference).⁹ First, DMKLM reported that the effect of performance on satisfaction was significantly different for the Japan/Korean comparison. The reanalysis suggests that this is an incorrect inference as the related path coefficient is invariant across the three datasets ($\chi^2 = 3.42, p > .10$). Second, DMKLM had concluded that the effect of role conflict on commitment was *not* significantly different for any of their pairwise comparisons. The reanalysis reveals that in actuality this effect does vary significantly for their cross-national data ($\chi^2 = 6.06, p < .05$). However, our reanalysis affirms DMKLM’s finding that the effect of role conflict on job satisfaction varies significantly ($\chi^2 = 10.92, p < .001$).

Based on the preceding results, a *combined* restricted model was estimated with all path coefficients restricted to be invariant across the cross-national datasets except for the two paths found to be significantly different in Table 2 (i.e., RC \longrightarrow JS, and RC \longrightarrow OC).¹⁰ This model yielded fit statistics as follows: $\chi^2 = 48.1, df = 21, p < .01, NFI = .95, NNFI = .96,$ and $CFI = .97$. Compared to the *unrestricted* model, this model was significantly poorer indicating that additional sources of cross-national differences were present ($\Delta\chi^2 = 34.0, \Delta df = 15, p < .01$). Based on multivariate test statistics, three additional paths were identified as discrepant.¹¹ Specifically, we found that the effect of (1) role ambiguity on job satisfaction was different in Japan, (2) role ambiguity on commitment was different in Korea, and (3) role conflict on role ambiguity was different in Japan, but invariant otherwise. A final restricted model was estimated yielding a $\chi^2 = 28, df = 18, p > .05, NFI = .97, NNFI = .98,$ and $CFI = .99$. Compared to the *unrestricted* model, this final model is *not* significantly different ($\Delta\chi^2 = 14, \Delta df = 12, p > .10$) indicating that the final restricted model is equivalent to the unrestricted model, at least in terms of model fit. The estimated parameters of this final model are in Table 3.

Taken together, the preceding reanalysis suggests that DMKLM’s analysis failed to uncover three sources of cross-national differences (i.e., RC \longrightarrow OC, RA \longrightarrow JS and RA \longrightarrow OC) and reported one spurious cross-national

difference (i.e., JP → JS). On the other hand, their findings concerning two cross-national differences (i.e., RC → JS and RC → RA) are upheld.

TABLE 2
Cross-National Differences in Estimated Path Coefficients^a

Variable Relationships	← DMKLM Analysis →			← Reanalysis →			
	← Bivariate Z-Values ^d →			Unrestricted Chi-Square ^b	Restricted Chi-Square ^c	Difference Chi-Square ^e	p-value ^f
	U.S./Japan	U.S./Korea	Japan/Korea				
RC → RA	3.26**	1.08	-1.89	14.07	18.89	4.82^g	> .05
RA → JP	-1.22	-.76	.33	14.07	14.79	.72	> .80
RC → JP	-.57	.28	.83	14.07	14.19	.12	> .90
RA → JS	1.52	.07	-1.29	14.07	19.39	5.32 ^g	> .05
RC → JS	-3.17**	-.75	2.06*	14.07	24.99	10.92***	< .001
JP → JS	-1.51	1.55	2.73**	14.07	17.49	3.42	> .10
RA → OC	.80	1.46	.76	14.07	19.95	5.88 ^g	> .05
RC → OC	.59	-.81	-1.36	14.07	20.13	6.06**	< .05
JS → OC	1.43	1.06	-.18	14.07	18.29	4.22	> .10

^a Statistically significant differences between DMKLM analysis and the reanalysis are in bold.

^b This is the *chi*-square corresponding to the model in which the path coefficients are allowed to vary across the three cross-national datasets. This is referred to as an unrestricted model and serves as a common baseline model for all comparisons. The corresponding degrees of freedom are 6.

^c This is the *chi*-square corresponding to the model in which the corresponding path coefficient is restricted to be *invariant* across the three cross-national datasets. This is referred to as a restricted model. The corresponding degrees of freedom are 8.

^d Reproduced from DMKLM's Table 4 [1992, p. 90].

^e This is a *chi*-square difference test. The *chi*-square statistic is obtained by computing the difference in *chi*-squares between the restricted and unrestricted models.

^f The degrees of freedom for this test equal 2.

^g Although the *chi*-square difference is nonsignificant, the multivariate Lagrange Multiplier (LM) test from an overall restricted model reveals that this path coefficient is significantly different in one of the three datasets. These differences are described in the text.

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

Differences in Estimated Path Coefficients

Overall model differences do not tell the whole story. This is because consistent overall results from two analyses may be obtained on account of dramatically disparate path coefficients (e.g., coefficients are consistently large and significant in one analysis, and *nonsignificant* in the second). This necessitates a closer comparative analysis of path coefficients (see Table 3). We focus on three issues: (1) explained variances (i.e., R^2) of dependent constructs, (2) magnitude and statistical significance of invariant coefficients, and (3) coefficients that vary cross-nationally.

Explained Variance. Table 3 reveals that the reanalysis yields higher levels of R -square. In some cases, the differences are striking. For instance, the R^2 for role ambiguity is .28 for Japanese data in the reanalysis and .13 in DMKLM's

– an increase of over 115%. For the Korean sample, DMKLM explain 48% of the variance in commitment compared to 71% in the reanalysis. Similar increases are evident across the board, with a minimum increase of 20%. These increases are nontrivial and stem mainly from DMKLM's neglect of measurement error.

Magnitude and Significance of Invariant Path Coefficients. In comparing studies, note that consistent support for the invariance of a given path (e.g., RC → RA) does not necessarily imply that the magnitude and significance of this path would also be equivalent across the two studies. Four such differences are especially noteworthy in Table 3. First, DMKLM had reported that the effect of role ambiguity on performance is invariant and estimated to be $-.34$ ($p < .001$), $-.20$ ($p < .01$), and $-.24$ ($p < .001$) in the U.S., Japanese and Korean data respectively. The reanalysis confirms DMKLM's conclusion of invariance. However, the estimated unstandardized coefficient is $-.51$ ($p < .001$) for each sample. Thus, DMKLM underestimate the magnitude of this relationship (e.g., $-.24$ vs $-.51$) and overestimate the variability in the estimated effects (i.e., from $-.20$ to $-.34$).

Second, DMKLM found that the effect of role conflict on performance is invariant and equals $.13$ (ns), $.19$ ($p < .01$) and $.09$ (ns) for the U.S., Japanese and Korean data respectively. The reanalysis reveals that this coefficient is $.35$ ($p < .001$) in each sample. Thus, while DMKLM had hypothesized that role conflict will have a negative effect on performance, the data show that this direct effect is positive, invariant and highly significant. This, however, does not imply that higher the role conflict, higher will be the performance of salespeople, as alluded to by DMKLM (p. 90). This is because the total effect of role conflict on performance is composed of a direct effect (role conflict → performance link) and an indirect effect as role conflict affects ambiguity, and ambiguity in turn affects performance. As such, the total effect of role conflict on performance is nonsignificant in each cross-national sample. Thus, in contrast to DMKLM, the reanalysis of their data shows that role conflict (1) has a significant, direct, positive effect, along with (2) a nonsignificant total effect on performance of salespeople, and (3) both these effects are invariant cross-nationally.

Third, in terms of the effect of role conflict on ambiguity, DMKLM found the coefficients to be $.63$, $.36$ and $.54$ (all $p < .001$) for the U.S., Japanese and Korean samples. DMKLM correctly attributed the lower effect for Japanese salespeople to an environment of "consensual building, harmony and free exchange of information." In reality, their data suggest that role conflict has an equal adverse impact in U.S. and Korea and its magnitude is $.73$ ($p < .001$). In Japan, the magnitude is as high as $.52$ ($p < .001$). This represents an increase of 44% for the Japanese data and, 35% and 16% for the Korean and U.S. data respectively. Thus, DMKLM underestimated this coefficient, notably

for the Japanese data. Moreover, DMKLM's report that, for the Japanese, the mean value of role conflict is the *highest* and significantly larger than for the U.S. and Korean data (their Table 2). Apparently, the operating environment of Japanese salespeople engenders significantly *greater* role conflict but its adverse effects on role ambiguity are less potent than in the U.S. and Korean samples.

TABLE 3
Estimated Path Coefficients for the Interrelationships
between Role Stressors and Work Outcomes

Dependent Variable	← DMKLM Analysis ^a →			← Reanalysis ^b →	
	Explanatory Variable	Standardized Coefficient	R ²	Unstandardized Coefficient	R ²
Sample 1: U.S. (N = 218)^c					
RA	RC	.63***	.40	.73***	.55
JP	RA	-.34***	.07	-.51***	.12
	RC	.13		.35***	
JS	RA	-.05	.44	.13	.58
	RC	-.58***		-.80***	
	JP	.20***		.27***	
OC	RA	-.15**	.61	-.20***	.73
	RC	-.15**		-.11	
	JS	.58***		.65***	
Sample 2: Japan (N = 220)^c					
RA	RC	.36***	.13	.52***	.28
JP	RA	-.20**	.05	-.51***	.19
	RC	.19**		.35***	
JS	RA	-.19**	.26	-.33**	.41
	RC	-.29***		-.24*	
	JP	.33***		.27***	
OC	RA	-.21***	.47	-.20***	.66
	RC	-.20***		-.11	
	JS	.47***		.65***	
Sample 3: Korea (N = 156)^c					
RA	RC	.54***	.29	.73***	.51
JP	RA	-.24***	.04	-.51***	.13
	RC	.09		.35***	
JS	RA	-.06	.29	.13	.45
	RC	-.50***		-.73***	
	JP	.06		.27***	
OC	RA	-.28***	.48	-.50***	.71
	RC	-.07		.22*	
	JS	.49***		.65***	

^aReproduced from DMKLM's Table 3 [1992, p. 89]. For the sake of comparison, we utilized the path coefficients for DMKLM's "full" model.

^bObtained from a simultaneous path analysis estimated across the three datasets with covariance matrix as input. Corrections for unreliability were applied in accord with the attenuation formula [Nunnally 1978, pp. 219-20].

^cDifferences between DMKLM and reanalysis coefficients that exceed .20 are highlighted. This amounts to a spread of about two standard errors since the average standard error for coefficients is about .10.

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$

Finally, DMKLM found that the effect of performance on satisfaction to be .20 ($p < .001$), .33 ($p < .001$) and .06 (ns) in the U.S., Japanese and Korean samples respectively. DMKLM had attributed the nonsignificant difference for the Korean sample to the noncontingent reward systems in Korea. The reanalysis shows that the unstandardized path coefficient equals .27 ($p < .001$) for each sample. Thus, DMKLM's inference of noncontingent reward systems in Korea appears to be in error.

Cross-National Differences in Path Coefficients. Two studies could also differ because they obtain a different pattern of cross-national differences. Three such differences are noteworthy in Table 3. First, DMKLM found that the effect of role conflict on satisfaction varies across nations. The reanalysis confirms this finding; however, DMKLM *underestimated* the coefficients in the U.S. and Korean samples ($-.58$ & $-.50$ vs. our $-.80$ & $-.73$), and *overestimated* it for the Japanese sample ($-.29$ vs. our $-.24$). Thus, as far as the effect of role conflict on satisfaction is concerned, the gap between the Japanese and U.S./Korean samples is larger than alluded to by DMKLM. Recall that the mean value of role conflict is *highest* for the Japanese sample. Despite this higher level of role conflict, its adverse effects on satisfaction are attenuated by a factor of one-third for Japanese salespeople.

Second, for the effect of role conflict on commitment, DMKLM concluded that the effects are invariant with estimated coefficients of $-.15$ ($p < .01$), $-.20$ ($p < .001$) and $-.07$ (ns) for the U.S., Japanese and Korean data respectively. In fact, the coefficients are $-.11$ (ns), $-.11$ (ns), and $.22$ ($p < .05$) for the U.S., Japanese and Korean data respectively. A scrutiny of the corresponding coefficients reveals dramatic shifts in the magnitude and significance of effects. DMKLM found that the effect of role conflict on commitment was significant for the U.S. sample; in fact this effect is nonsignificant. For the Korean sample, in contrast, opposite results emerged.¹²

Finally, the reanalysis shows that the effect of role ambiguity on satisfaction is significantly larger in the Japanese sample ($-.33$, $p < .001$ vs. DMKLM's $-.19$) than either in the U.S. (.13 vs. DMKLM's $-.05$) or the Korean (.13 vs. DMKLM's $-.06$) samples. As such, the adverse effects of role ambiguity are significantly larger for the Japanese sample. Surprisingly, DMKLM's results show that the Japanese have the highest mean level of role ambiguity. Likewise, the reanalysis shows that the effect of role ambiguity on commitment is significantly larger for the Korean sample ($-.50$, $p < .001$ vs. DMKLM's $-.28$) than either the U.S. ($-.20$ vs. DMKLM's $-.15$) or the Japanese ($-.20$ vs. DMKLM's $-.21$) samples. Thus, the adverse effects of role ambiguity on commitment are more critical for the Korean salespeople and, consistent with this, they have the lowest mean value for organizational commitment.

DISCUSSION AND IMPLICATIONS

Although we have utilized DMKLM's study to highlight the issues involved, the aim of this paper has been to draw the attention of cross-national researchers to (1) four specific measurement issues that are relevant to much empirical research, (2) the resulting inferential pitfalls and substantive errors that are likely to go unnoticed if these issues are ignored, and (3) provide methodological approaches for tackling these measurement concerns. Specifically, we discussed problems and issues involved in utilizing standardized versus unstandardized coefficients, accounting for measurement error and unequal reliability, controlling overall error rate via simultaneous analysis, and construct equivalence assessment. Previous research has paid little attention to either the measurement concerns or their implications in cross-national studies. Yet, the reanalysis of DMKLM's data demonstrates that, when these issues are ignored, inferential errors increase considerably, exhibit complex patterns, and their magnitude and direction are unpredictable.

In all, the reanalysis reveals that DMKLM appear to have inadvertently (1) committed one Type-I error (i.e., inferring a specious cross-national difference), (2) committed three Type-II errors (i.e., overlooking a significant cross-national difference), (3) *underestimated* all of the twelve explained variances (i.e., R^2), and (4) provided numerous instances of *over* and *under*-estimation of the magnitude and significance of specific path coefficients. The net effect is that DMKLM's and the reanalysis yield disparate substantive outcomes about cross-national differences. For instance, in sharp contrast to DMKLM, the reanalysis shows that the most intriguing cross-national differences involve role conflict and ambiguity for the Japanese sample. The Japanese experience the *highest*, and significantly greater role conflict and ambiguity than their U.S. and Korean counterparts. Yet the adverse effect of role conflict on satisfaction is the *smallest*, and reduced by a factor of one-third for the Japanese relative to U.S. and Korean salespeople. Similarly, the adverse effect of role ambiguity on satisfaction is the *highest*, and amplified threefold for the Japanese relative to U.S. and Korean salespeople. Evidently, the Japanese are able to cope with high levels of role conflict, but *not* role ambiguity. DMKLM's results do not unravel such intriguing insights.

While these differences are not trivial, it would be inappropriate to view our research simply as a critical rejoinder to DMKLM's study. Rather, our research should be viewed as a call to all cross-national researchers to pay increasing attention to measurement issues or risk serious inferential errors. A decade ago, Davis et al. [1981] had cautioned that in some cross-national studies, "what might appear to be cross-national difference could turn out to be solely a reflection of variations in the reliability of the underlying measurements employed in the analysis." We heighten this caution and note that unequal reliability is but *one* of several measurement issues facing a cross-

national researcher. The combined interactive effect of these measurement issues is unpredictably complex resulting in some cross-national differences showing up as spuriously significant while others appear erroneously insignificant, and the substantive havoc caused by under and overestimation of the magnitude of effects is unmistakably pervasive. We conclude that a lack of attention to measurement issues interferes in our ability to draw valid inferences and, for this reason, measurement issues warrant serious consideration in future cross-national research.

NOTES

1. For instance, the transformational index for the coefficient of X_1 on Y_1 in sample 1 is (s_{x11}/s_{y11}) , where s_{x11} and s_{y11} are the standard deviations of X_1 and Y_1 in sample 1. Thus, the standardized coefficient $s\beta_{11}$ equals $\beta_{11} (s_{x11}/s_{y11})$. Readers will note that the standardized coefficient ($s\beta$) will equal the unstandardized coefficient (β) if, and only if, for *each* sample X_1 , X_2 and Y have a mean of zero and standard deviation of 1.
2. One sufficient condition is that X_1 and X_2 are uncorrelated in each sample.
3. This is valid as long as equation (1) is an unbiased regression equation for the relation between X_1 and Y .
4. For independent tests, the overall error rate is given by $[1 - (1 - \alpha)^{27}]$ where α equals .05 for a 95% confidence level.
5. The "adjustment" formula is as follows. If the reliability of a given construct in a given sample is α , the measurement error is fixed at $(1 - \alpha)\sigma^2$ where σ is the standard deviation of the construct.
6. The "*chi*-square difference" test is based on the notion that the difference between two *chi*-square statistics is itself distributed as a *chi*-square. The test statistic is the mathematical difference between the *chi*-square for the "fully restricted" and the "unrestricted" models, with the degrees of freedom computed as the corresponding difference in the degrees of freedom of the two models.
7. A covariance matrix can be computed from the correlation matrix and standard deviations of individual constructs. However, the use of covariance matrix does not allow tests for construct equivalence. Raw data were not available. The EQS software is available from BMDP [Bentler 1989] and is suited for LVSE analysis. An alternative option is LISREL.
8. For a model that fits the data well, the *chi*-square test should be *nonsignificant*. For the models estimated, the *chi*-square test is significant at $p = .05$. However, this test is sensitive to sample size and is prone to Type-I errors when sample sizes are "high" (say, > 200). The fit indices are less sensitive to sample size. As noted earlier, fit indices > .95 indicate a reasonably well-fitting model [Bollen 1989].
9. One reviewer pointed out that these conclusions are not necessarily correct because DMKLM conducted a bivariate test whereas our reanalysis utilized a multivariate test. Our test is multivariate because all *three* path coefficients are restricted to be equal in the restricted model under the invariant hypothesis. The alternative hypothesis is that at least one pair of these coefficients is significantly different. By contrast, DMKLM examine each pair of coefficients separately (each at .05) with the alternative hypothesis that the coefficients are not equal to each other. As noted under our discussion of "*overall error rate*," multiple bivariate comparisons are likely to inflate overall error rate resulting in erroneous inferences. A multivariate test helps control overall error rate.
10. However, the RC \rightarrow OC path was set to be invariant across U.S. and Japan data, but different in Korea. We thank a reviewer for pointing this out.

11. The additional paths were as follows: (1) RA \longrightarrow JS was found to be different in Japan, while in U.S. and Korea it was invariant ($\Delta\chi^2 = 9.3$), (2) RA \longrightarrow OC was different in Korea but invariant in U.S. and Japan ($\Delta\chi^2 = 5.2$), and (3) RC \longrightarrow RA was different in Japan, while being invariant in U.S. and Korea ($\Delta\chi^2 = 4.8$). These paths were identified by using the multivariate Lagrange Multiplier (LM) test (Bentler 1989). Only the RC \longrightarrow RA difference was reported by DMKLM.

12. We depict the *direct* effect of RC on OC. RC also affects OC indirectly because of RC's direct effects on RA, JP and JS. The total effect of RC on OC is invariant as it equals $-.60$, $-.58$ and $-.54$ in the U.S., Japanese and Korean data.

REFERENCES

- Adler, Nancy. 1983. A typology of management studies involving culture. *Journal of International Business Studies*, 14(3): 29–47.
- Alwin, Duane F. 1988. Measurement and the interpretation of effects in linear structural equation models. In S. Long, editor, *Common problems/Proper solutions*. Beverly Hills, Calif.: Sage Publications.
- Bentler, Peter. 1989. *EQS: Structural equations program manual*. Los Angeles, Calif.: BMDP Statistical Software.
- Bollen, Kenneth A. 1989. *Structural equations with latent variables*. New York: John Wiley.
- Brislin, Richard W. 1976. Comparative research methodology: Cross cultural studies. *International Journal of Psychology*, 11: 215–29.
- Byrne, Barbara. 1989. Multigroup comparisons and assumption of equivalent construct validity across groups: Methodological and substantive issues. *Multivariate Behavioral Research*, 24: 503–23.
- Cole, David A. & Scott E. Maxwell. 1985. Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach. *Multivariate Behavioral Research*, 20: 389–417.
- Davis, Harry L., Susan P. Douglas & Alvin J. Silk. 1981. Measure unreliability: A hidden threat to cross-national marketing research. *Journal of Marketing*, 45(2): 98–109.
- Drasgow, Fritz & Ruth Kanfer. 1985. Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70: 662–80.
- Dubinsky Alan J., Ronald E. Michaels, Masaaki Kotabe, Chaelin Lim & Hee-Chal Moon. 1992. Influence of role stress on industrial salespeople's work outcomes in the United States, Japan, and Korea. *Journal of International Business Studies*, 23(1): 77–99.
- Green, Robert T. & Philip D. White. 1976. Methodological considerations in cross-national consumer research. *Journal of International Business Studies*, 7 (Fall–Winter): 81–87.
- Hummel, T. J. & J. Sligo. 1971. Empirical comparison of univariate and multivariate analysis of variance procedures. *Psychological Bulletin*, 76: 49–57.
- Jöreskog, Karl G. 1971. Simultaneous factor analysis in several populations. *Psychometrika*, 36: 409–26.
- Kahn, R. L., Donald M. Wolfe, R. Quinn, J.D. Snoek & R. H. Rosenthal. 1964. *Organizational stress*. New York: John Wiley.
- Klockars, Alan & Gilbert Sax. 1986. *Multiple comparisons*. Beverly Hills, Calif.: Sage Publications.
- Mullen, Michael R. 1995. Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies*, 26(3): 573–596.

- Nunnally, Jum. 1978. *Psychometric theory*. New York: McGraw Hill.
- Parameswaran, Ravi & Attila Yaprak. 1987. A cross-national comparison of consumer research measures. *Journal of International Business Studies*, 8(1): 35-49.
- Sekaran, Uma. 1983. Methodological and theoretical issues and advancements in cross-cultural research. *Journal of International Business Studies*, 14(2): 61-73.
- Triandis, Harry C. 1980. *Handbook of cross-cultural psychology*. Boston: Allyn and Bacon.
- Walker, Orville C., Jr., Gilbert A. Churchill, Jr. & Neil M. Ford. 1977. Motivation and performance in industrial selling: Present knowledge and needed research. *Journal of Marketing Research*, 14 (May): 156-68.