

**TESTING MEASUREMENT EQUIVALENCE
IN CROSS-NATIONAL RESEARCH:
AN EMPIRICAL TEST ACROSS U.S. AND JAPAN**

by

B. M. BENSAOU*
M. Coyne**
and
N. VENKATRAMAN†

97/41/ABA/TM

- * Associate Professor of Information Technology and Management at INSEAD, Boulevard de Constance, Fontainebleau 77305 Cedex, France.
- ** PhD Candidate at Boston University, School of Management, 704 Commonwealth Avenue, Boston MA 02215, USA.
- † Professor of Management at Boston University, School of Management, 704 Commonwealth Avenue, Boston MA 02215, USA.

A working paper in the INSEAD Working Paper Series is intended as a means whereby a faculty researcher's thoughts and findings may be communicated to interested readers. The paper should be considered preliminary in nature and may require revision.

**Testing Measurement Equivalence in Cross-National Research:
an Empirical Test across U.S. and Japan**

M. Bensaou

INSEAD
Boulevard de Constance
77309 Fontainebleau Cedex, France
(33-1) 6072-4021

Michael Coyne

and

N. Venkatraman

Boston University
School of Management
704 Commonwealth Avenue
Boston MA 02215

We thank INSEAD, the Euro-Asia Centre, and Boston University Systems Research center for research support in preparation of this paper.

**Testing Measurement Equivalence in Cross-National Research:
an Empirical Test across U.S. and Japan**

Abstract

We propose a conceptual and analytical framework for assessing measurement equivalence in cross-national strategy research. We illustrate the use of this diagnosis procedure as we test the equivalence of the measurement models previously used across the U.S. and Japan to capture eight conceptual constructs derived from transaction cost economics (e.g., asset specificity, reciprocal investments) and organization theory (e.g., conflict, trust, task characteristics). Results of the multi-group LISREL analysis conducted on a dataset of 447 relationships reveal some constructs with strong cross-national equivalence, others with need for content respecification and some for re-conceptualization. We derive implications and suggestions about how to conduct empirical strategy research in cross-national settings.

INTRODUCTION

Under intensifying international competition the business environment is becoming global. Not only managers in small and large firms but subsequently strategy and management researchers as well are increasingly interested in better understanding the similarities and differences between countries. Their concerns range from understanding how to negotiate in a cross-cultural context, how to manage foreign direct investment (Kogut, 1983), joint ventures or strategic alliances with foreign partners (Kogut, 1988, Hamel, et al., 1989), how to structure and manage multinationals (e.g., Bartlett and Ghoshal, 1989; Doz, 1976; Prahalad, 1976; Hedlund, 1986). How do multinationals balance the pressures for both local isomorphism and parent central influence to leverage corporate capabilities across national borders (e.g. Ghoshal and Nohria, 1989; Rosenzweig and Nohria, 1994; Westney, 1993, see review by Earley and Singh, 1995).

In addition, with the recent success of competitors with a different cultural background, an emerging body of work in strategy concerns a better understanding of the boundary conditions of Anglo-Saxon practice and its underlying theories of management. Japanese firms, for instance, operate within a different cultural and institutional context (Dore, 1987; Gerlach, 1992; Nishiguchi, 1994), have developed different conceptions of strategy (Hamel, 1989; Prahalad and Hamel, 1990) and operate different organizational arrangements (Nonaka and Takeuchi, 1995; Aoki, 1988). Then, 'is it meaningful to develop models that fail to generalize outside of the United States (Earley and Singh, 1995 p. 329)?'

Collectively the strategy and management field has been moving up Adler's typology of cross-national research studies (Adler, 1983) -- shifting from parochial, ethnocentric or polycentric studies to comparative, geocentric and synergistic types of research approaches. Yet, just as these studies are trying to achieve an understanding of how management and

organizational phenomena relate to cultural and national characteristics, they typically give little attention to some commensurate methodological issues (with notable exceptions such as Janssens, et al., 1995 or Murtha, et al., 1996). As we face some ambiguities regarding the universal applications of our theoretical concepts we also face the mirror issues about the generalizability of the corresponding measurements we readily administer across different national settings (Kotabe and Murray, 1990).

This is a particularly important issue as theories and measures cannot and should not be de-coupled (Bagozzi and Phillips, 1982; Ember et al., 1991; Mullen, 1995; Sekaran, 1983; Singh, 1995). We believe it has become critical for strategy researchers conducting their studies in a cross-national setting to systematically consider the issues of whether their constructs, samples and final measures are equivalent across the compared national populations (Mayer, 1978; Sekaran, 1983; Douglas and Craig, 1983; Parameswaran and Yaprak, 1987; Bhalla and Lin, 1987; Aulakh and Kotabe, 1993). Inadequate testing for the invariance of data across national groups weakens the interpretations that can be derived from cross-national empirical research.

This paper focuses on these significant yet customarily disregarded issues of measurement equivalence and the important question as to whether the empirical similarities and differences uncovered do reflect a real phenomenon which will inform theory or are merely artifacts of the measurement instrument. We seek to contribute to strategy research by proposing and illustrating the use of a diagnosis procedure researchers can regularly utilize to assess measurement equivalence across national samples before testing their substantive hypotheses. The purpose of this paper is not a test of nomological validity which requires development of specific causal relationships between constructs but this is solely a demonstration of a simple approach we used to assess metric equivalence (i.e., an ingredient

of measurement equivalence) for eight constructs regularly used in transaction cost economics and organization theory research to test hypothesis across national settings.

The paper is organized in six sections. The first one reviews the concept of measure equivalence and defines metric equivalence. In the following section we build upon existing literature in marketing, strategy and international management and define a diagnosis procedure to assess metric equivalence. The third section describes the study by Bensaou (1992) which provided the cross-national data we re-analyzed to explain by example the use of the diagnosis procedure. We specifically examine how well key issues of measurement were treated before data collection during the early stages of research design, sampling and instrument development. The fourth 'results' section describes the analysis we conducted on the data he collected in the U.S. and Japanese automobile industries. Finally we discuss the implications of these findings and conclude with suggestions for greater systematic attention paid to measurement equivalence in cross-national strategy and management research.

MEASUREMENT EQUIVALENCE

Functional, conceptual, sampling and translation equivalencies

Equivalence in content and quality is needed to 'ensure that empirical findings reflect similarities and differences between countries, rather than the spurious effect of sociocultural differences in response to a research instrument, administration procedures, or lack of adaptation of the research design and plan to a specific sociocultural environment' (Douglas, and Craig, 1983 p. 132). To generate data that are comparable from one country to another we need to examine the equivalence of various aspects of the data collection process. Douglas and Craig (1983) distinguish between different types of equivalence. First, are the constructs and concepts studied equivalent (i.e., construct equivalence)? 'Functional equivalence is concerned

with the role of objects and behavior in society from a macrocultural level, while conceptual equivalence is concerned with the interpretation that individuals place on objects, stimuli, or behaviors and whether these are expressed in similar ways in different countries' (Craig and Douglas 1983, p. 138).

Secondly, are the observed measures of the conceptual constructs equivalent across the two groups (i.e., measurement equivalence). For this, we first need to assess whether the sampling procedures and resulting samples collected in each country are equivalent. Translation equivalence is a critical issue for construct validity, since this is the stage at which the latent construct is defined in operational terms for informants from each country. These four methodological threats to validity need to be addressed before primary data collection, as early as research design, operationalization, instrument development and sampling stages. Finally, once data is collected strategy and management researchers conducting cross-national studies need to test for equivalence of the measures used and assess whether the initial care and efforts paid off. This paper concerns precisely this latter issue of metric equivalence and proposes then illustrates the use of a diagnosis procedure for a statistical test.

Metric equivalence

Metric equivalence deals with the issue of whether the same measurement model holds across different, distinct and exclusive populations, such as national samples. Does the score obtained on a scale in one country have the same meaning and interpretation in another country? Recent advances in structural equation modeling, in particular with respect to measurement, make it possible to analyze and compare measurement models across distinct groups (Fornell, 1987; Bollen, 1989; Jöreskog and Sörbom, 1989; also see Bagozzi for an early application in consumer behavior research across groups). In particular, LISREL provides multi-group analysis procedures to test several hierarchical hypotheses that

correspond to increasing degrees of metric equivalence. Following Bagozzi (1983) and Bollen (1989) we define a conceptual construct as metric equivalent across countries when it has the same (a) dimensionality, (b) measurement pattern, and (c) error variance.

At one extreme some measures have universal meaning while other have meaning that is unique to a specific country. While one might refer to this as context validity, we use the term metric equivalence to avoid possible confusion with external validity. There is a clear distinction between equivalence of measurement and external validity, although the two are somewhat related. Equivalence of measurement focuses on the relationship between the theoretical constructs and variables across distinct contexts, while external validity focuses on the ability to generalize the results of hypothesis testing (i.e., usually between constructs) to populations of interest. Singh (1995) also reminds that assessing metric equivalence is ‘not an all-or-nothing’ answer. Testing for it can be used to identify ‘offending scale items’ that violate metric equivalence expectations in spite of the attention given to construct and sampling equivalence prior to data collection. Metric equivalence assessment can indicate which items in the measurement instrument should be eliminated from the statistical analysis so that substantive conclusions can be drawn and identify those items that need further development in future research.

ANALYTICAL APPROACH

Specification of the model

Assume an observed variable x and its measurement model. Using LISREL notation (Jöreskog and Sörbom, 1989) the general model can be expressed as a set of structural equations of the form:

$$x = \Lambda_x \xi + \delta$$

Matrix Λ_x refers to the factor loadings matrix relating the measures x to their corresponding theoretical latent construct ξ , while δ refers to the vector of error terms - errors of measurement or measure-specific components. The measurement model for variable x in the U.S. and Japan samples are therefore defined by the following parameter matrices:

- Λ_x^{USA} , Λ_x^{Japan} (matrices of factor loadings), and

- Θ_x^{USA} , Θ_x^{Japan} (covariance matrices of δ)

As Bollen (1989, p. 356) states 'comparability or invariance in models represent a continuum.' He suggests a hierarchy of invariance (or equivalence) that can be assessed along two overlapping dimensions: one is model form and the other is similarity in the parameter values (see also Mullen, 1995). The first level examines whether variable x loads on the same number of dimensions or factors in the two national samples:

- $H_{n=k}$: same form (same dimensions)

If there is equivalence in model form, we can further test for the second level of 'factorial' equivalence (Singh 1995, p. 604), i.e., whether the loadings (λ) linking the latent variable x to the observed variable ξ are the same in the two country samples:

- H_Λ : $\Lambda_x^{USA} = \Lambda_x^{Japan}$

Finally, if factorial equivalence is achieved we can test the third and higher form of equivalence with the equality of measurement error variances (δ):

- $H_{\Lambda\theta}$: $\Lambda_x^{USA} = \Lambda_x^{Japan}$ and Θ_x^{USA} and Θ_x^{Japan}

With LISREL procedures parameters in each of these matrices can be set fixed, free or constrained. If we set no constraints across the U.S. and Japan samples, each sample can be analyzed separately. However, if we set constraints across the two samples, we can test any form of invariance across the two samples. The relative fit of these three restricted models, nested in less constrained models, can be done using chi-square difference tests.

Testing the model

Figure 1 illustrates our analytical framework. We use multi-sample LISREL analysis (Jöreskog and Sörbom, 1989 p. 227) as it is applicable to the holistic construal based on the concept of true scores (Lord and Novick, 1968; Bagozzi and Phillips, 1982). Our analytical approach follows two key steps. First, before testing for equivalence across the U.S. and Japan samples, we begin by assessing the unidimensionality of each of the constructs of interest on the pooled dataset. As will appear in the analysis, this test may result in re-specifying the original measurement model based on a theoretical re-interpretation of the measures and a better relative fit as compared to the original model initially used to construct the instrument. In a second step, we actually test for metric equivalence or country-specificity of the re-specified model.

Figure 1 about here

We suggest beginning, when appropriate, with a single factor model with all questionnaire items loading on a single latent construct. Based on these results (i.e., overall and relative fit of the model and the pattern of standardized residuals) and eventually a re-examination of the content of the items we identify the model to be tested for equivalence across countries. For instance, we found poor fit for the single factor model for the 'asset

specificity' construct as measured by Bensaou (1992) when all five original indicators load on one latent variable. Factor analysis identified a two factor model with two indicators each which we labeled 'knowledge specificity' and 'process specificity' after theoretical re-interpretation. Hence, we dropped item 1 for asset specificity (see Appendix A) and assigned items 2 and 3 to 'knowledge specificity' and items 4 and 5 to 'process specificity'.

Once we identify a re-specified model with good relative fit, we can proceed to test for metric equivalence and use the diagnosis procedures developed above and summarized in figure 1. For each of the three hypotheses, we use the χ^2 statistic (or $\Delta\chi^2$ asymptotically distributed as χ^2) as our test (Bentler and Bonett, 1980; Bagozzi, 1983; Anderson and Gerbing, 1988). These sequential chi-square difference tests are asymptotically independent (Steiger et al., 1985).

The first hypothesis $H_{n=k}$ states that the construct, e.g., 'knowledge specificity', has the same number of dimensions or factors in the two country samples. In this instance, our test determines that the construct of 'knowledge specificity' has two factors both in the U.S. and Japan. To test $H_{n=2}$ Jöreskog and Sörbom (1989) suggest the following procedure: to fix the scale for ξ_1 and ξ_2 , fix one of the λ parameters equal to one for each factor across both countries. The other λ elements, along with the diagonal elements of Θ_δ and the covariance matrix Φ are free in both countries. There are no equality constraints imposed in this problem. The model only states that the number of factors is the same for both countries. Λ_x for the second context (i.e., Japan) is specified to have the same pattern and starting values as for the first context (i.e., the US). The overall value of χ^2 is the sum of the χ^2 that would be obtained if the two contexts were analyzed separately. In other words, this χ^2 is a measure of the

overall fit of the model in both countries. In cases when $H_{n=k}$ does not hold we stop the procedure. The measures are not equivalent across samples.

If the hypothesis is confirmed, we proceed to the next step. The second hypothesis H_{Λ} states that the coefficients linking the latent construct (e.g., ‘knowledge specificity’) to the observed indicators are the same in both national samples. The model specification is essentially the same as for $H_{n=2}$, except for the addition of the constraint of Λ_x as invariant. We test this hypothesis with $\Delta\chi^2$, a measure of relative fit (Bagozzi, 1983, Anderson, 1987). This is a departure from Jöreskog and Sörbom (1989) who use absolute fit in the test of each of these hypotheses. Bagozzi (1983), following Bentler and Bonett (1980), equates the use of the relative fit (i.e., the incremental fit index) as an indication of practical relevance as opposed to statistical relevance. Relative fit is the preferred measure of fit because it captures only the difference based on the added constraint for each subsequent problem. Our use of $\Delta\chi^2$ tests on increasingly constrained nested models is also consistent with Anderson and Gerbing (1988).

If we find invariance in the item loadings on each factor of the construct, the next step is to test hypothesis $H_{\Lambda\theta}$. This final hypothesis states that the error terms are similar in the U.S. and Japan samples. For example, is the error the same for ‘knowledge specificity’ indicators in Japan and the USA? Again, we test the hypothesis with $\Delta\chi^2$. To test $H_{\Lambda\theta}$, the specification is essentially the same as for H_{Λ} , except for the addition of the constraint of Θ_x as invariant. This establishes a relative test for equality in the error variances. Equality in scaling is generally a higher priority than equality of measurement error variances (Bollen,

1989), so H_{Δ} precedes hypothesis $H_{\Delta\theta}$. Figure 1 and table 1 below summarize the proposed diagnosis procedure.

Table1 about here

THE EMPIRICAL STUDY

To demonstrate the use and usefulness of this framework for assessing metric equivalence across countries we used primary data Bensaou (1992) collected for his study of the strategic processes in buyer-supplier relationships in the U.S. and Japanese automobile industries. Supplier relations have received much attention since the early 1980's with direct comparison between the way U.S. and European firms manage their supply chain management as compared to the Japanese approach. Transaction cost analysis as well as organization theory concepts have been used to address the question of the degree of vertical integration, make-or-buy decision, arm's length relationships and relational contracting based on trust and long term commitment. From the variables available in the original dataset we selected eight variables we anticipated to be 'international-sensitive', and yet anchored in mainstream strategy research as demonstrates table 2: 'asset specificity' and 'reciprocal investments' (from transaction cost analysis), 'trust', 'conflict' and 'reputation' (from political economy), 'task characteristics' and 'information amount' (from organization theory and information processing view) -- for details about the initial study and the substantive questions examined see Bensaou (1992, 1997) and Bensaou and Venkatraman (1995).

These theoretical disciplines, primarily developed in an Anglo-Saxon tradition, have influenced many empirical studies conducted as multi-country comparative studies as well as single-country study in non Anglo-Saxon settings such as Japan or Korea. Typically, these

studies employ common definition of the constructs across countries and rarely test whether these concepts are equivalent and whether the measures employed to test their hypotheses are equivalent. In particular, the growth in strategy research involving Japan as an environment for market entry or Japanese companies as suppliers, customers or competitors warrants greater systematic attention to construct and measurement equivalence. In particular, one important question is whether theories developed in an Anglo-Saxon intellectual tradition, with their underlying assumptions and values would equally operate in Japan. What is their degree of generalizability? What are the boundaries conditions?

Table 2 about here

In this paper, as a first step, we seek to propose greater systematic attention to measurement issues and the risk they pose to the reliability and validity of the comparative results. Specifically, do the survey items (for quantitative studies) or interview questions (for qualitative studies) administered to American and Japanese managers and their firms measure equivalently underlying concepts drawn from a common theoretical tradition, and increasingly applied across national settings, e.g., asset specificity, trust or conflict. Table 2 provides a summary of the constructs Bensaou (1992) measured in both national settings, together with their use in prior research (of which some is empirically cross national). Appendix A provides the list of items for each of these constructs.

This paper proposes a procedure to test metric equivalence once the data has been collected, yet in the following paragraphs we need to first establish how the other equivalence issues related to cross-national empirical validation were addressed before and during data collection. As Douglas and Craig (1983) indicate the general concern is to establish comparability between data collected in different cultural contexts. 'We should thrive for data

that have, as far as possible, the same meaning or interpretation, and the same level of accuracy, precision of measurement, or reliability across national settings’.

Functional, conceptual and category equivalence. The U.S. and Japan are two developed countries, well beyond their industrial revolution and in fact leading the way into the information revolution, both have a long history of industrial and manufacturing development activities, and both benefit from a developed, mature and prosperous automobile sector. We therefore expect high levels of functional equivalence, i.e., the key ‘objects’ (e.g., buyers, suppliers, engineers, purchasers, components) and behaviors (e.g., trust, conflict) to have similar role in society at the macrocultural level. We expect these to be even more so similar within the same technological context of designing, producing and assembling cars.

Conceptual equivalence focuses on the individual level variations in attitude and behavior rather than on societal and cultural norms. Are objects and behaviors expressed and interpreted in similar ways in different countries? Exploratory fieldwork was conducted in both countries in the local language and by the same researcher (as recommended by Sekaran and Martin 1982), and did not reveal any gap in the way engineers and purchasing agents interviewed in Japan or the U.S. expressed or interpreted the key concepts, such as ‘specificity of assets’, ‘amount of information’ or ‘trust’. This should not come as a surprise as the study did not deal with psychological or individual level construction but rather with ‘technical’ issues shared by practitioners well informed about different practices around the globe. The globalization of the auto industry and the resulting wide exchange of information and best practice has resulted in framing the issues in converging ways around the world and using the same language and categories to talk about the drivers of performance or quality for instance. We therefore accept conceptual equivalence across the U.S. and Japanese samples collected in the automobile industry.

Translation and sampling equivalence. Once construct equivalence was verified, we examined the issue of measurement equivalence, including translation, sampling and finally metric equivalence. In both countries target informants were used to receiving research questionnaires (though in Japan more often from industry associations or ministry research groups than university professors conducting independent academic research as in the U.S.) and were used to 7-point Likert scales. To reduce threats to translation equivalence, the same bilingual researcher developed and refined the measurement instrument in parallel in the two languages, conducting interviews in the Detroit area with U.S. informants employed in U.S. firms and with Japanese informants employed in the Japanese transplant factories. Subsequently, each of the resulting versions of the questionnaire was further translated back into the other language and submitted for review to industry and academic experts (e.g., Nobeoka, Fujimoto and Takeichi for the Japanese version; Cusumano and Clark for the English version).

Sampling equivalence was achieved by developing and implementing the same sampling procedure (as recommended by Sekaran 1983) in all participating firms so as to have the similar relevant individuals respond to the questionnaire in each country. In addition, both samples are representative of each country's automobile industries as all car assemblers participated. To ensure representativeness at the supplier and component level and avoid selection bias, the researcher submitted to the key contact in each company a stratified list of 50 components (established with auto experts) from which he or she had to select a product first and then send the questionnaire to the key informant directly in charge of the account. The sample for the study resulted in $n = 447$ questionnaires, with $n = 140$ in the U.S. and $n = 307$ in Japan. Each questionnaire was administered to a manager responsible for a distinct car component sourced from a distinct supplier and who would inform about characteristics of the component and its market, characteristics of the supplier (e.g., reputation), characteristics

of the buyer-supplier relationship (e.g., asset specificity, reciprocal investments, commitment, trust, conflict) and finally characteristics of his or her job (e.g., information amount, task characteristics).

RESULTS

Once data was collected we tested for metric equivalence using the framework developed above. Do the scores obtained in the U.S. sample have the same meaning and interpretation in the Japanese context? As figure 1 explains, we test each of the eight constructs first for unidimensionality with the pooled sample and then for metric equivalence across the two national samples using sometimes a re-specified model. A tabular summary of results is presented in Table 3 which includes statistics for χ^2 , p-value, and $\Delta\chi^2$ as they apply to the three assessment objectives discussed in Table 1.

Table 3 about here

Asset Specificity

Consistent with transaction costs analysis we define 'asset specificity' as the extent to which some identified investments are specific to a relationship. Implicit is the fact that these are difficult or expensive to transfer to another relationship or may lose the value when redeployed to another party. The fieldwork executed in the auto companies based in the U.S. and Japan revealed four areas of investments critical to a car company that can be deployed exclusively into a supplier. Special tools and dies are built or lent specifically for the component delivered by a given supplier. The customization can lead to tailoring some of the design work to some key components (such as air conditioning or wire harnessing). Clearly the time and effort spent by purchasing and engineering staff to learn the supplier's business

practices constitute investments that cannot be transferred to another relation. The same holds for the time spend on developing the relationship. This is time and effort that is not spent developing new business opportunities with another supplier and the benefits of these initiatives accrue primarily to this relationship.

The initial one-factor four-indicator measurement model for ‘asset specificity’ has poor absolute fit $\chi^2(2)=38.67$ ($p=.000$) and significant standardized residuals (maximum magnitude = 5.60). The model needs re-specification. Informed by the pattern of the residuals and how they match the content of the question items we re-specified the model along two dimensions: ‘physical specificity’; and ‘knowledge specificity’ with two indicators each¹. The re-specified two-factor two-indicator model displays acceptable relative fit [$\chi^2(1)=7.78$ ($p=.005$); $\Delta\chi^2(1)=30.89$ ($p<.005$), GFI = .990, AGFI=.902].

We then used the hierarchy of hypotheses displayed in table 1 above as the basis for testing metric equivalence. The first hypothesis, $H_{n=2}$, tests for form equivalence, i.e., whether ‘asset specificity’ indeed has two factors in the U.S. and Japanese samples. Table 3 shows the results. The resulting $\chi^2 =5.55$ with 2 degrees of freedom (df) is insignificant ($p=.062$). We therefore fail to reject $H_{n=2}$ and conclude that the measurement model effectively captures two dimensions in the two datasets. We can therefore proceed on testing the subsequent hypotheses. Hypothesis, H_{λ} , tests whether factor loadings are invariant simultaneously for both latent variables ‘knowledge specificity’ and ‘process specificity’ across U.S. and Japanese data. We also fail to reject this second hypothesis H_{λ} : [$\Delta\chi^2(2)=3.53$ ($p>.1$)] and conclude that measurement items 1 and 2 (see appendix A) load on ‘physical specificity’ (and items 3 and 4 load on ‘knowledge specificity’) in a similar pattern in the U.S.

¹ See Appendix. ‘Physical specificity’ is measured by ‘asset specificity’ items 1 and 2. ‘Knowledge specificity’

and Japanese samples. We can test the third hypothesis, $H_{\Lambda\Theta}$, for equality of measurement error variance across the two country samples. We fail to reject this hypothesis $H_{\Lambda\Theta}$: [$\Delta\chi^2(4)= 5.94$ ($p>.1$)] which signifies an excellent fit of the model to the data from both countries. These findings provide us with strong evidence that the re-specified measurement model for 'asset specificity' is the same in the U.S. and Japanese samples.

Reciprocal Investments

Reciprocal investments represent those specific investments made by the supplier into the relationship and which act as dependence-balancing mechanism (Heide and John 1988; Beier and Stern 1969; Emerson 1962). Fieldwork conducted in the two national contexts revealed five areas where suppliers may place customer specific assets: equipment, technical knowledge, design, manufacturing and managerial capabilities.

The initial one-factor five-indicator model for 'reciprocal investments' (see appendix A) had poor fit and large standardized residuals. After inspection of the content of the items we decided to re-specify the model. We tested two different sets of models for absolute and relative fit and ultimately identified a two-factor model with two indicators each. Conceptual examination of the content of the items suggests one factor refers to 'investments in knowledge' and a second one refers to 'process investments'. The re-specified model has good absolute fit: $\chi^2=.02$ with one degree of freedom (df), goodness of fit index (GFI)=1.000 and adjusted goodness of fit index (AGFI)=1.00. The standardized residuals between items is not significant (maximum magnitude =.136).

Subsequently, we accept the first hypothesis for form equivalence, $H_{n=2}$ [$\chi^2(2)=.07$ ($p= .963$)] and conclude that the 'reciprocal investments' construct effectively has two

dimensions in the two country data sets. We further accept the second hypothesis for factor loadings equivalence, $H_{\Lambda} : [\Delta\chi^2(2)=.27 (p>.1)]$ and conclude that items 2 and 3 load on ‘investments in knowledge’ (similarly items 4 and 5 load on ‘process investments’) in a similar pattern in the U.S. and Japanese samples. Finally, we accept the third hypothesis of error variance equivalence $H_{\Lambda\Theta} [\Delta\chi^2(2)=.27 (p>.1)]$ and conclude that the measurement error for each of the indicators is similar across Japanese and U.S. data. We are therefore in a strong position to accept measurement equivalence across the U.S. and Japan (hence measurement and construct equivalence) for the ‘reciprocal investments’ construct as measured by our re-specified model.

Trust

While, ‘asset specificity’ and ‘reciprocal investments’ are mutual investments specific to the relationship whereby each member holds the other hostage (Anderson 1985, Heide and John 1990), trust in contrast contributes to the reduction of uncertainty about the opportunistic behavior by the other partner (Axelrod 1984, Dore 1983, Ouchi 1980). Interviews revealed that sharing sensitive information and compatible management approaches reflect high levels of trust.

The one-factor, three-indicator measurement model for ‘trust’ has good absolute fit: $\chi^2(1)= 0.00$, goodness of fit index (GFI)=1.000 and adjusted goodness of fit index (AGFI)=1.00. The standardized residuals between items are not significant (maximum magnitude =.058). The results for the three metric measurement tests lead us to accept all three hypotheses: { $H_{n=1} : [\chi^2(1)=.19 (p=.665)]$; $H_{\Lambda} : [\Delta\chi^2(2)=.08 (p>.1)]$; $H_{\Lambda\Theta} : [\Delta\chi^2(3)=5.48 (p>.1)]$ }. Our measurement instrument for ‘trust’ therefore proves equivalent across the two national samples.

Supplier Reputation

A supplier's reputation for consistent performance and for respecting its engagements contributes to reducing the buyer's uncertainty about the future actions of its partner. Transaction cost reasoning would suggest lower search costs for suppliers with better reputation. The one-factor, four- indicator measurement model for 'supplier reputation' has poor absolute fit $\chi^2(2)=11.71$ ($p=.003$) and significant standardized residuals (maximum magnitude = 3.36). Model re-specification leads to a two-factor two-indicator model: 'supplier's reputation for outcomes' and 'supplier's reputation for process'². The re-specified model has acceptable fit [$\chi^2(1)=1.58$ ($p=.209$); $\Delta\chi^2(1)=10.13$ ($p<.005$); GFI=.998; AGFI=.982]. Furthermore, the standardized residuals between items is not significant (maximum magnitude =1.25).

The tests for form equivalence and factor loading equivalence are conclusive indicating that the two-dimension re-specified model holds across both countries and that the individual measures have the same pattern of loading on these two dimensions across the two groups. Indeed, we fail to reject $H_{n=2}$: [$\chi^2(2)=2.05$ ($p=.358$)] and H_{Λ} : [$\Delta\chi^2(2)=.76$ ($p>.1$)]. However, the test for measurement error reveals inconclusive. We find a poor fit of this model to the data and reject $H_{\Lambda\Theta}$: [$\Delta\chi^2(4)=21.82$ ($p<.005$)].

Our measurement model for 'supplier reputation' is therefore not operating in the same way in the data from the U.S. and Japanese populations and is not a valid instrument for cross-cultural comparisons between these two populations. Yet, another interpretation, based on invariance in number of factors and item loadings, would suggest that the factor

² See Appendix. 'Supplier's reputation for outcome' is measured by items 1 and 4. 'Supplier's reputation for process' is measured by items 2 and 3.

structure of our measurement model is acceptable but the items need to be further refined so that they are equally interpretable across the U.S. and Japan.

Commitment

Commitment reflects the socio-political context (Arndt 1983; Benson 1975; Gardner and Cooper 1988) underlying the relationship, and ranges along a cooperative-conflictual continuum, and directly affects the extent to which information, for instance, is freely exchanged between the buyer and the supplier because or in spite of the governance of the relationship (Reve and Stern, 1984). Fieldwork associated higher commitment with sharing of responsibilities, risks, burden and benefits in the relationship.

The one-factor four-indicator measurement model for 'commitment' has poor absolute fit: [$\chi^2(2)=5.82$ ($p=.054$), GFI=.993, AGFI=.966]. However, the standardized residuals between items are not significant (maximum magnitude =2.29). The pattern of the residuals and the content of the items however do not suggest re-specification. We therefore decided to diagnose for metric equivalence using the initial measurement model, without accepting or rejecting a claim of unidimensionality.

We fail to reject the first hypothesis for form equivalence $H_{n=1}$ [$\chi^2(4)=4.57$ ($p=.335$)] and conclude that the 'commitment' latent variable effectively has one dimension in the two country data sets. Proceeding to the second hypothesis for factor loadings equivalence, we fail to reject H_{λ} [$\Delta\chi^2(3)=4.81$ ($p>.1$)] and conclude that the four items used (see appendix A) load on 'commitment' in a similar pattern in the U.S. and Japanese samples. Finally, we reject the third hypothesis of error variance equivalence $H_{\lambda\theta}$ [$\Delta\chi^2(4)=43.55$ ($p<.005$)] and conclude that the measurement error for each of the indicators is not equal across Japanese and U.S. data. Our measurement model for 'commitment' is not 'metric' equivalent across the two

countries. Yet, rejection of this third hypothesis only suggests that, like for ‘supplier reputation’, our model may simply need improved item content to enhance consistency across countries. Just as Singh (1995) reminds us ‘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.’ (p. 606).

Task Characteristics

Uncertainty is seen as a key determinant of the design of buyer-supplier relation (Bensaou and Venkatraman (1996) building on Galbraith (1973, 1977), Tushman and Nadler (1978), March and Simon (1958) and Weick (1979)). A major source of uncertainty is the task and its underlying level of analyzability and variety. We hence started with a two-factor measurement model, which proved to have poor absolute fit [$\chi^2(1)=4.34$ ($p=.037$), GFI=.995, AGFI=.952], but acceptable magnitude for standardized residuals (maximum magnitude =2.08). We therefore used this initial model to test for metric equivalence.

We then applied our diagnostic framework and tested for the first hypothesis $H_{n=2}$ for same form. The $\chi^2 = 2.88$ with 2 degrees of freedom is insignificant ($p=.237$). This insignificant χ^2 , and resulting acceptance of $H_{n=2}$, suggests that our measurement model for ‘task characteristics’ has the same number of dimensions across the two national populations. As figure 1 indicates we can proceed to testing the next hypothesis. For H_{Λ} , the χ^2 difference $\Delta\chi^2=0.18$ with 2 degrees of freedom is insignificant ($p>.1$). Again, we can accept H_L and

move on to test the third hypothesis of error variance equality. The resulting χ^2 difference $\Delta\chi^2(4)=61.91$ with 4 degrees of freedom is highly significant ($p<.005$), providing evidence that the additional invariance constraint on error terms has a bad effect on model fit. Hypothesis $H_{\Lambda\Theta}$ must be rejected³, indicating that we cannot conclude that our measurement model is equivalent across Japan and the US. Yet, rejection of only the third most stringent hypothesis of invariance in error variance suggests that we may simply need improved item content to enhance consistency across countries.

Information Amount

Information processing needs arise from uncertainty and should be matched by the information processing capabilities inherent to a relationship. As such they determine the design of the buyer-supplier relation. We started our analysis including the two dimensions Bensaou (1992) had captured in his survey, i.e., information amount and information ambiguity. However, tests of the two-factor six-indicator model revealed poor absolute fit [$\chi^2(8)=26.65$ ($p=.001$), GFI=.981, AGFI=.950]. After examination of the content of the items we dropped items 5 and 6 for 'information ambiguity' from the model. We then tested a one-factor four-indicator model for 'information amount' which displayed a poor absolute fit [$\chi^2(2)=11.67$ ($p=.003$), GFI=.987, AGFI=.934] and significant standardized residuals between items (maximum magnitude =3.40). Further re-examination of the content of the questions suggests re-specifying the model as a two-factor model: 'monitoring amount' (items 2 and 4) and 'exchange amount' (items 1 and 3) with two indicators each. The derived two-factor four-indicator measurement model has acceptable absolute and relative fit [$\chi^2(1)=.29$

³ The finding on $H_{\Lambda\Theta}$ is based on an intermediate solution. The model was found non-admissible after 350 iterations. Θ_{δ} was not positive definite.

($p=.592$), $GFI=1.000$, $AGFI=.997$, $\Delta\chi^2(1)=11.38$ ($p<.005$)] and the pattern of the residuals is consistent with item content.

Just as for ‘supplier reputation, ‘commitment’ and ‘task characteristics’ we conclude same number of dimensions and equal measure loading across the two groups. Indeed, we fail to reject the first two hypothesis in our diagnosis procedure: $H_{n=2}$ [$\chi^2(2)=.59$ ($p=.743$)] and H_{Λ} [$\Delta\chi^2(2)=2.22$ ($p>.1$)]. However, our re-specified model fails to satisfy the third requirement as we reject $H_{\Lambda\Theta}$ [$\Delta\chi^2(4)=50.24$ ($p<.005$)]. ‘Information amount’ as we defined and measured it satisfies some degree of metric equivalence, but may need further ‘content’ improvement to reach full equivalence. We, however, know that the model used already achieves the ‘same form’, and ‘factorial loading equivalence’ (i.e., each item has the same loading (within statistical bounds) and on the same factor across the U.S. and Japanese data).

Conflict

Conflict also reflects the socio-political context within which the relationship operates and affects its information processing capabilities. Interviews with buyer and supplier informants indicated five areas of typical disagreement: price, design, quality, delivery schedules and inventory policies.

The initial one-factor, five-indicator measurement model, however, displays a poor fit to the pooled data and large standardized residuals (maximum magnitude = 8.56). This negative result together with a re-examination of the content of the items suggests model re-specification. We tested two alternative models for comparative fit. The resulting model is a two-factor four-item measurement model (we dropped item 1) with two underlying

dimensions: ‘outcome conflict’ and ‘process conflict’ with two indicators each⁴. The absolute fit and the relative fit of this re-specified model prove good [$\chi^2(1)=.56$ ($p=.454$), $GFI=.999$, $AGFI=.993$, $\Delta\chi^2(1)=5.92$ ($p<.025$)] and the standardized residuals between items are not significant (maximum magnitude =.748).

We then used the hierarchy of hypotheses displayed in figure 1 as the basis for testing metric equivalence. The first hypothesis, $H_{n=2}$, proves conclusive and indicates that the measurement model used in the Japanese and U.S. questionnaires indeed captures some related construct with two dimensions or factors. The $\chi^2=2.59$ with 2 degrees of freedom (df) is insignificant ($p=.274$). We can therefore proceed on testing the subsequent hypotheses. The second hypothesis, H_{Λ} , which tests whether factor loadings are invariant simultaneously for both latent variables ‘outcome conflict’ and ‘process conflict’ across U.S. and Japanese data. We reject this second hypothesis [$\Delta\chi^2(2)=6.02$ ($p>.1$)] Though not necessary, we also tested for the third hypothesis and rejected it as well: $H_{\Lambda\Theta}$ [$\Delta\chi^2(4)=88.82$ ($p<.005$)]. The items we used to measure conflict do load onto two latent conceptual variables, but differ in terms of their loadings and error variance across the two U.S. and Japanese data samples.

In summary, failure to accept these last two hypotheses suggests that our measurement model of conflict is not metric equivalent across cultural contexts, specifically with respect to the pattern of loadings and error variance. There seems to be indeed two underlying factors for ‘conflict’ captured by our instrument, but they may have different interpretations in the two countries.

⁴ See Appendix for items measuring conflict. ‘Outcome conflict’ is measured by items 2 and 3. ‘Process conflict’ is measured by items 4 and 5.

DISCUSSION

Table 3 provides a summary of the results for the eight constructs examined empirically. We find our measurement models fully metric equivalent for 'asset specificity', 'trust' and 'reciprocal investments'. We found four models to be partially equivalent, i.e., they passed the 'same form' and 'same pattern of loadings' test but failed the most stringent 'same error variance' hurdle. These four models may need further refinement of the items content. Finally, we found that our proposed measurement model for 'conflict' only passed the same form hypothesis test. The items seem to be tapping two dimensions in both countries but their meaning and interpretation differs across the two populations.

The results for the three first constructs in table 3 give us some assurance that these measurement schemes could be used in the study of interorganizational relationships between U.S. and Japan. This may help generalize this operationalization to other nations as well, given that the proposed diagnosis procedure is again applied after data is collected. The main implication is that the comparative results obtained with these higher quality constructs and measures are more likely to reflect 'real' substantive effects and demonstrate strong nomological validity for the theoretical relationships being tested.

The second set of results reveals four constructs that are equivalent across the U.S. and Japanese population in terms of their dimensions and the way the indicators loaded on the latent factors, yet there were differences in the variability in the way the questions were answered (i.e., differences in measurement error) across the two groups. The wording for some or all of the items used for these constructs may allow differing variance in interpretation across contexts. In such case, careful refinement of the items may be sufficient to achieve full metric equivalence. Specifically, we suggest for greater care during translation of the items. Yet, we believe these scales can be used to test substantive relationships

between conceptual constructs, but the data cannot be pooled. The 'etic' approach should prevail. The diagnosis procedure we propose has the potential, if used systematically and replicated on the same scales in different countries by different researchers, to advance our survey and construct designs. Reported violating items and scales can be removed from further use and equivalent ones can be safely administered.

The final set of findings about 'conflict' raises insightful questions and challenges for strategy and management researchers. Though we cannot determine whether the failure of metric equivalence is due to the 'conceptualization' of conflict in the first place or its 'measurement' in the two countries. In the first case, we need to further explore, in the tradition of the 'emic' approach, i.e., the underlying definition of conflict, its conceptual antecedents and consequences in each culture. In other words, we may accept that for our purposes an attitudinal and behavioral concept such as 'conflict' is unique to the U.S. or Japanese culture and is best understood in each culture's own terms. The measures will therefore be specifically adapted to each national sample. Our re-specified model for 'conflict' can therefore be accepted within the framework of this approach. However, 'this implies making inferences about cross-national differences and similarities in qualitative or judgmental terms, since each measure is culture-specific' (Douglas and Craig 1983 p. 133). On the other hand, this same finding can be interpreted within the 'etic' perspective. This approach is primarily concerned with identifying and assessing universal attitudinal and behavioral concepts and developing pan-cultural or 'culture-free' concepts (Elder, 1976). Our model then fundamentally requires different items to tap the two dimensions we already 'uncovered' to achieve factorial and error variance equivalence.

CONCLUSIONS

As firms internationalize and researchers further apply concepts and theories, typically embedded in one cultural background, to different national samples it is becoming critical for the strategy field to address the issue of comparability, in particular of data equivalence. Do U.S. theories and their associated measurement models (indeed, each measurement model has an implicit theory) apply equally in Japan? 'To the extent that each country has certain idiosyncratic features, and is characterized by unique sociocultural behavior patterns and values, attitudinal and behavioral phenomena may be expressed in unique ways. Specifically, insofar as there is commonality between countries, comparable constructs and concepts can be identified' (Douglas p. 133).

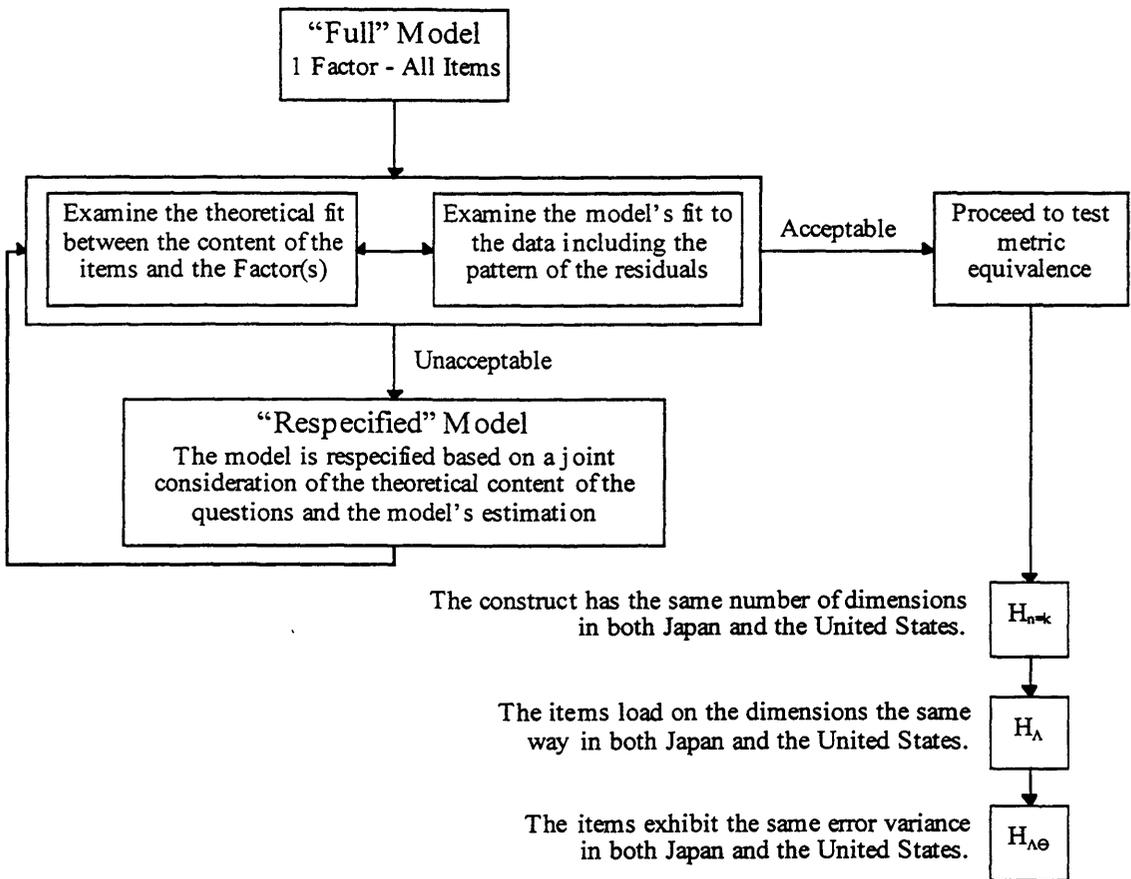
The diagnosis framework we propose provides both substantive and methodological insight. At a practical level the procedure can be used even in a single country test for 'metric equivalence' across different independent groups of respondents. This could be, for instance, across industry sectors for multi-industry studies, across gender, age groups, or informant roles. Beyond the mere validation of equivalence, the procedure can advance the 'interpretability' of the concepts behind the measurement models.

Cross-national strategy research using the diagnosis procedure can incrementally contribute a pool of 'metric equivalent' constructs and their measurement models for common use. Yet, we believe the procedure can provide even greater contribution in those cases where one or more of the hypotheses is rejected. Immediate rejection of the first hypothesis implies eliminating the scale from further analysis of substantive hypotheses. But, more importantly it raises a substantive question about the concept and the underlying theory. For instance, the results reported above alert us that 'conflict' may mean different things in the U.S. and in Japan. This may call for re-visiting our conceptualization of conflict as a universalistic concept in the 'etic' tradition or further adaptation of the measures to each culture in the 'emic' tradition. For the models violating the less dramatic 'error variance' equivalence only,

one remedy can be to use the scales as such, yet recognizing and interpreting the results in the light of failed test. The other necessary 'ex-post' remedy is to further improve the scales with a tighter content specification of the items, in particular during translation. But it does not imply reconceptualization.

Overall, we hope that this paper can contribute to raise awareness about important measurements issues in cross-national strategy research and provide a diagnosis procedure strategy researchers can apply when developing and testing empirical models. We believe that the next major breakthrough in strategy will occur through greater attention accorded to generalization of results across different countries and regions; such generalizations are to be firmly rooted in systematic attention to measurement issues and challenges.

Figure 1: Conceptual and Analytical Framework



Source: Adapted and extended from Bagozzi (1983), Anderson (1987), Jöreskog and Sörbom (1989).

Table 1: Hypothesis Testing for Country-specificity

<i>Assessment Objective</i>	<i>Hypothesis</i>	<i>Measurement Concept</i>
Does the construct have the same dimensions in both Japan and the United States?	$H_{n=k}$	Posits equality in the number of factors across contexts.
Do the items load on the dimensions the same way in both the Japan and the United States?	H_{Λ}	Posits equality in the pattern of measurements across contexts. The test is based on relative fit as compared to the model for $H_{n=k}$.
Do the items exhibit the same error in both Japan and the United States?	$H_{\Lambda\Theta}$	Posits equality in both measurement parameters and error variances. This is a test of the entire measurement model across contexts (Bagozzi 1983). However, we test based on relative fit as compared to the model for H_{Λ} when the additional constraint of invariant error-variances is added.

Table 2: Constructs: Theoretical Anchors and Items

Construct	Theoretical Anchors	Exemplar(s)
Asset Specificity	This is a central concept to transaction cost economics (TCE) and has been used in every study examining TCE and interorganizational relationships.	Williamson, 1975 Heide, et al., 1988 Anderson, 1985
Trust	Trust reflects the intersection of organizational theory and organizational economics and has been widely used as an important determinant of channel structure.	Williamson, 1985 Ouchi, 1980 Macneil, 1980
Supplier Reputation	Supplier Reputation essentially reflects the history of the supplier's interactions with its customers.	Anderson and Weitz, 1989, 1992
Reciprocal Investment	Reciprocal Investments are viewed as "the extent to which the principal invests resources in the relationship with the agency" (Zaheer and Venkatraman 1994). They restrain the opportunistic actions of partners, hence protecting the parties specific investments (Heide & John 1990)	Pfeffer and Salancik, 1978 Heide, et al., 1988, Heide and John, 1990 Zaheer and Venkatramam, 1995
Conflict	Conflict and Commitment reflect organizational theory concepts, specifically as socio-political processes (Arndt 1983, Benson 1975) that have been argued to impact free information sharing in dyads (Reve and Stern 1984).	
	Conflict is a situation in which one party perceives the other to be engaged in a behavior that is preventing or impeding it from achieving its goals.	Reve and Stern, 1976 Gaski, 1984 Lusch, 1976
Commitment	Commitment embodies concepts from TCE and organizational theory. The "essence of commitment is stability and sacrifice ... Committed relationships entail confidence that the affiliation will last, a desire to make them last, and a willingness to make sacrifices to grow and maintain the alliance. Other terms used to describe these relationships include 'partnerships' or 'alliances' -- terms suggesting unity or fusion" (Anderson and Weitz 1992).	Gardner and Cooper, 1988 Henderson, 1990 Anderson & Weitz, 1992
Task	Task is a general concept which has the dimensions of Task Analyzability and Task Variety. These have been examined in organizational theory research.	Perrow, 1967
Information Amount	Information is a general construct which has the dimensions of Information Monitoring Amount and Information Exchange Amount. These have been examined in the measurement branch of institutional economics and information processing view.	Alchian and Demsetz, 1972

Table 3: Results

Level of equivalence	Construct	$H_{n=k}$		H_{Λ}			$H_{\Lambda\Theta}$		
		χ^2 (df)	p-value	χ^2 (df)	p-value	$\Delta\chi^2$ (df)	χ^2 (df)	p-value	$\Delta\chi^2$ (df)
Metric equivalence	Asset Specificity	5.55 (2)	0.062	9.08 (4)	0.059	3.53 (2)	15.02 (8)	0.059	5.94 (4)
	Reciprocal Investments	.07 (2)	0.963	.34 (4)	0.987	.27 (2)	.84 (8)	1	.50 (4)
	Trust	.19 (1)	0.665	.27 (3)	0.996	.08 (2)	5.51 (5)	0.357	5.48 (3)
Factorial equivalence & same form equivalence	Supplier Reputation	2.05 (2)	0.358	2.81 (4)	0.591	.76 (2)	24.63 (8)	0.002	21.82 (4) ($<.005$)
	Commitment	4.57 (4)	0.335	9.38 (7)	0.227	4.81 (3)	52.94 (11)	0.000	43.55 (4) ($<.005$)
	Task characteristics	2.88 (2)	0.237	3.06 (4)	0.549	0.18 (2)	64.97 (8)	0.000	61.91 (4) ($<.005$)
	Information Amount	.59 (2)	0.743	2.81 (4)	0.59	2.22 (2)	53.05 (8)	0.000	50.24 (4) ($<.005$)
Same form equivalence only	Conflict	2.59 (2)	0.274	8.61 (4)	0.071	6.02 (2) ($<.05$)	97.43 (8)	0.000	88.82 (4) ($<.005$)

Appendix A - Measurement Items

Asset Specificity

To what extent do you believe your firm has made major investments specifically for its relationship with this supplier?

	STRONGLY DISAGREE		NEITHER AGREE NOR DISAGREE			STRONGLY AGREE	
Your firm has made significant investments in tooling dedicated to its relationship with this supplier	1	2	3	4	5	6	7
Some of your firm's products have been tailored to using this component bought from this supplier	1	2	3	4	5	6	7
It has taken your firm a lot of time and effort to learn this supplier business practices to become effective	1	2	3	4	5	6	7
Your firm has spent a lot of time and effort to develop and nurture the relationship with this supplier	1	2	3	4	5	6	7

Reciprocal Investments

In your opinion, does the production of this component require capabilities and labor skills that are relatively unique to THIS SUPPLIER, or can this component be produced with standard capabilities and labor skills by any supplier?

	VERY STANDARD					VERY UNIQUE TO THIS SUPPLIER	
Layout, facilities & tooling	1	2	3	4	5	6	7
Technological knowledge	1	2	3	4	5	6	7
Design skills and capabilities	1	2	3	4	5	6	7
Manufacturing skills and capabilities	1	2	3	4	5	6	7
Managerial skills and experience	1	2	3	4	5	6	7

Trust

How would you describe the degree of mutual trust between your firm and this supplier?

Extremely weak 1 2 3 4 5 6 7 Extremely strong

How comfortable do you feel about sharing very sensitive information with this supplier?

Very uncomfortable 1 2 3 4 5 6 7 Very comfortable

Would you say that your firm and this supplier have compatible management philosophies?

Very incompatible 1 2 3 4 5 6 7 Very compatible

Supplier Reputation

How much do you agree or disagree with the following statements about this supplier's *reputation* and *track record* in the industry?

	STRONGLY DISAGREE		NEITHER AGREE NOR DISAGREE			STRONGLY AGREE	
This supplier has an excellent reputation for holding to its commitments	1	2	3	4	5	6	7
This supplier has an excellent track record for responding to design changes	1	2	3	4	5	6	7

Your firm can rely on this supplier to meet spec. requirements without detailed supervision	1	2	3	4	5	6	7
---	---	---	---	---	---	---	---

This supplier has an excellent reputation as a low cost provider	1	2	3	4	5	6	7
--	---	---	---	---	---	---	---

Commitment

Would you say that your firm and this supplier equally share the burden and the benefits of the relationship?

	Your firm has more of the share		Equally shared with this supplier			This supplier has more of the share	
Risks	1	2	3	4	5	6	7
Burden	1	2	3	4	5	6	7
Benefits	1	2	3	4	5	6	7
Responsibility for component quality	1	2	3	4	5	6	7

Task Characteristics

Please indicate how much you agree or disagree with the following statements about the work you do with this supplier for this component.

	STRONGLY DISAGREE			NEITHER AGREE NOR DISAGREE			STRONGLY AGREE
There is clearly a known way to do your job when it relates to this supplier (e.g. a manual)	1	2	3	4	5	6	7
There are established practices & procedures you follow in doing your job with this supplier	1	2	3	4	5	6	7
You basically perform repetitive tasks	1	2	3	4	5	6	7
You do the same tasks in the same way most of the time	1	2	3	4	5	6	7

Information Amount and Ambiguity

Please indicate the extent to which you agree with the following statements:

	STRONGLY DISAGREE			NEITHER AGREE NOR DISAGREE			STRONGLY AGREE
You need to continuously gather a great deal of information about other suppliers and their products	1	2	3	4	5	6	7
You need to closely monitor this suppliers behavior	1	2	3	4	5	6	7
You need to exchange information with this supplier on a very timely and urgent basis	1	2	3	4	5	6	7
You need to exchange very detailed information with this supplier on a continuous basis	1	2	3	4	5	6	7

There are many ambiguous events and issues in your dealings with this supplier for this component 1 2 3 4 5 6 7

Most issues with this supplier are well understood, and face-to-face discussions are not so important 1 2 3 4 5 6 7

Conflict

Please indicate the degree to which each of the following categories cause disagreements between your firm and this supplier:

	MINIMAL DISAGREEMENT		SOME DISAGREEMENT			EXTREME DISAGREEMENT	
Component price	1	2	3	4	5	6	7
Component design	1	2	3	4	5	6	7
Quality levels	1	2	3	4	5	6	7
Delivery schedules	1	2	3	4	5	6	7
Inventory policies	1	2	3	4	5	6	7

References

- Adler, N. J. (1983). 'A Typology of Management Studies Involving Culture', *Journal of International Business Studies*, **Fall**, pp. 29-47.
- Alchian, A. A. and H. Demsetz (1972). 'Production, Information Costs, and Economic Organization', *American Economic Review*, **62 (December)**, pp. 777-795.
- Anderson, E. (1985). 'The Salesperson as Outside Agent or Employee: A Transaction Cost Analysis', *Marketing Science*, **4 (3)**, pp. 234-254.
- Anderson, E. and B. Weitz (1992). 'The Use of Pledges to Build And Sustain Commitment in Distribution Channels', *Journal of Marketing Research*, **24**, pp. 18-34.
- Anderson, E. and B. Weitz (1989). 'Determinants of Continuity in Conventional Industrial Channel Dyads', *Marketing Science*, **8 (4)**, pp. 310-323.
- Anderson, J. C. (1987). 'An Approach for Confirmatory Measurement and Structural Equation Modeling of Organizational Properties', *Management Science*, **33(4)**, pp. 525-541.
- Anderson, J. C. and D. W. Gerbing (1988). 'Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach', *Psychological Bulletin*, **103(3)**, pp. 411-423.
- Aoki, M. (1988). *Information, Incentives, and Bargaining in the Japanese Economy*. Cambridge University Press, Cambridge.
- Aulakh, P. S. and M. Kotabe (1993). 'An Assessment of Theoretical and Methodological Developments in International Marketing', *Journal of International Marketing*, **1(2)**, pp. 5-28.
- Bagozzi, R. P. (1983). 'A Holistic Methodology for Modeling Consumer Response to Innovation', *Operations Research*, **31(1)**, pp. 128-176.
- Bagozzi, R. P. and L. W. Phillips (1982). 'Representing and Testing Organizational Theories: a Holistic Construal', *Administrative Science Quarterly*, **29 (3)**, pp. 459-489.
- Bhalla, G. And L. Lin (1987). 'Cross-Cultural Marketing Research: A Discussion of Equivalence Issues and Measurement Strategies', *Psychology & Marketing*, **4(4)**, pp. 275-285.
- Bartlett, C. A. and S. Ghoshal (1989). *Managing Across Borders: The Transnational Solution*. Harvard Business School Press, Boston, MA.
- Bensaou, M. (1997). 'Interorganizational cooperation: the role of information technology. An empirical study of U.S. and Japanese supplier relations', *Information Systems Research*, **forthcoming**.
- Bensaou, M. and N. Venkatraman (1995). 'Configurations of Interorganizational Relationships: A Comparison between US and Japanese Automakers', *Management Science*, **41 (9)**, pp. 1471-1492.
- Bentler, P. M. and D. G. Bonett (1980). 'Significance Tests and Goodness-of-fit in the Analysis of Covariance Structures', *Psychological Bulletin*, **88**, pp. 588-600.

- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons, New York.
- Dore, R. (1987). *Taking Japan Seriously: A Confucian Perspective on Leading Economic Issues*. Stanford University Press, Stanford, CA.
- Douglas, S. P. and C. S. Craig, (1983). *International Marketing Research*. Prentice Hall, Englewood Cliffs, NJ.
- Earley, C. and H. Singh (1995). 'International and Intercultural management research: What's next?', *Academy of Management Journal*, **38(2)**, pp. 327-340.
- Elder, J. (1976). 'Comparative cross-national methodology', *Annual Review of Sociology*, **2**.
- Ember, C. R., M.H. Ross, M.L. Burton and C. Bradley (1991). 'Problems of Measurement in Cross-Cultural Research Using Secondary Data', *Behavior Science Research*, **25(4)**, pp. 187-216.
- Fornell, C. (1987). 'A Second Generation of Multivariate Analysis: Classification of Methods and Implications for Marketing Research'. In M. J. Houston (eds.), *Review of Marketing 1987*, American Marketing Association, Chicago, pp. 407-450.
- Gardner, J. and M. C. Cooper (1988). *Elements of Strategic Partnership*. Logistics Resources Inc.
- Gaski, J. F. (1984). 'The Theory of Power and Conflict in Channels of Distribution', *Journal of Marketing*, **48**, pp. 9-29.
- Gerlach, M. (1992). *Alliance Capitalism: the social organization of Japanese business*. University of California Press, Berkeley.
- Ghoshal, S. and N. Nohria (1989). 'Internal Differentiation within Multinational Corporations', *Strategic Management Journal*, **10**, pp. 323-337.
- Hamel, G. a. P., C.K. (1989). 'Strategic Intent', **May-June**, pp. 63-76.
- Hamel, G., Y. Doz and C. K. Prahalad (1989). 'Collaborate with Your Competitors - and Win', *Harvard Business Review*, **Jan-Feb**, pp. 133-139.
- Hedlund, G. (1986). 'The Hypermodern MNC - A Heterarchy ?', *Human Resource Management*, **25(1)**, pp. 9-35.
- Heide, J. B. and G. John (1990). 'Alliances in Industrial Purchasing: The Determinants of Joint Action in Buyer-Supplier Relationships', **27 (February)**, pp. 24-36.
- Heide, J. B. and G. John (1988). 'The Role of Dependence Balancing in Safeguarding Transaction-Specific Assets in Conventional Channels', *Journal of Marketing*, **52 (January)**, pp. 20-35.
- Henderson, J. C. (1990). 'Plugging into Strategic Partnerships: The Critical I/S Connection', *Sloan Management Review*, **31(3)**, pp. 7-18.
- Janssens, M., J. M. Brett and F. J. Smith (1995). 'Confirmatory cross-cultural research: testing the viability of a corporation-wide safety policy', *Academy of Management Journal*, **38(2)**, pp. 364-382.
- Jöreskog, K. G. and D. Sörbom (1989). *LISREL VII User's Reference Guide*. Jöreskog and Sörbom/SPSS Inc., Chicago, IL.

- Kogut, B. (1988). 'Joint Ventures: Theoretical and Empirical Perspectives', *Strategic Management Journal*, **9**, pp. 319-332.
- Kogut, B. (1983). 'Foreign direct investment as a sequential process'. In C.P. Kindleberger and D. Autretsch (eds.), *The multinational corporation in the 1980s*. MIT Press, Cambridge, MA, pp. 35 - 56.
- Kotabe, M. and J. Y. Murray (1990). 'Linking Product and Process Innovation and Modes of International Sourcing in Global Competition: A Case of Foreign Multinational Firms', *Journal of International Business Studies*, **21(3)**, pp. 383- 425.
- Lord, F. and M. Novick (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA.
- Lusch, R. F. (1976). 'Sources of Power: Their Impact on Intrachannel Conflict', *Journal of Marketing Research*, **13**, pp. 382-390.
- Macneil, I. R. (1980). *The New Social Contract, An Inquiry into Modern Contractual Relations*. Yale University Press, New Haven, CT.
- Mayer, C. S. (1978). 'Multinational Marketing Research: The Magnifying Glass of Methodological Problems', *European Research*, **6**(March), pp. 77-83.
- Mullen, M. R. (1995). 'Diagnosing Measurement Equivalence in Cross-National Research', *Journal of International Business Studies*, **26(3)**, pp. 573- 596.
- Murtha, T., S. A. Lenway and R. P. Bagozzi (1996). 'Global Mindsets and Cognitive Shift in a Complex MNC', *Strategic Management Journal*, **forthcoming**.
- Nishiguchi, T. (1994). *Strategic Industrial Sourcing: The Japanese Advantage*. Oxford University Press, New York.
- Nonaka, I. and H. Takeuchi (1995). *The Knowledge-Creating Company*. Oxford University Press, Oxford.
- Ouchi, W. G. (1980). 'Markets, Bureaucracies, and Clans', *Administrative Science Quarterly*, **25**, pp. 129-141.
- Parameswaran, R. and A. Yaprak (1987). 'A Cross-Cultural Comparison of Consumer Research Measures', *Journal of International Business Studies*, **8(1)**, pp. 35-49.
- Perrow, C. (1967). 'A Framework for the Comparative Analysis of Organizations', *American Sociological Review*.
- Pfeffer, J. and G. R. Salancik (1978). *The External Control of Organizations: A Resource Dependence Perspective*. Harper & Row, New York.
- Prahalad, C. K. and G. Hamel (1990). 'The Core Competence of the Corporation', *Harvard Business Review*, **May-June**, pp. 79-91.
- Reve, T. and L. W. Stern (1976). 'The Relationship between Interorganizational Form, Transaction Climate, and Economic Performance in Vertical Interfirm Dyads'. In L. a. S. R. E. Pellegrini (eds.), *Marketing Channels -- Relationships and Performance*.
- Rosenzweig, P. and N. Nohria (1994). 'Influences of human resource management practices in multinational firms', *Journal of International Business Studies*, **20(2)**, pp. 229-252.
- Sekaran, U. (1983). 'Methodological and Theoretical Issues and Advancements in Cross-Cultural Research', *Journal of International Business Studies*, **Fall**, pp. 61-73.

- Singh, J. (1995). 'Measurement Issues in Cross-National Research', *Journal of International Business Studies*, **26(3)**, pp. 597-617.
- Steiger, J.H., A. Shapiro, A. and W. M. Browne (1985). 'On the Multivariate Asymptotic Distribution of Sequential Chi-squared Statistics', *Psychometrika*, **43**, pp. 225-243.
- Westney, D. E. (1993). 'Institutional Theory and the multinational corporation'. In S. G. a. D. E. Westney (eds.), *Organization theory and the mutinational corporation*, St. Martin's Press, New York, pp. 53-76.
- Williamson, O. E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press, New York, NY.
- Williamson, O. E. (1985). *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. Free Press, New York.
- Zaheer, A. and N. Venkatramam (1995). 'Relational Governance as an Interorganizational Strategy: An Empirical Test of the Role of Trust in Economic Change', *Strategic Management Journal*, **16**(June), pp. 373-392.