



Assessing the cross-national invariance of formative measures: Guidelines for international business researchers

Adamantios
Diamantopoulos¹ and
Nicolas Papadopoulos²

¹Department of Business Administration,
University of Vienna, Vienna, Austria; ²Sprott
School of Business, Carleton University, Ottawa,
Canada

Correspondence:

A Diamantopoulos, Department of Business
Administration, University of Vienna,
Bruenner Strasse 72, Vienna, A-1210,
Austria.
Tel: +43 1 4277 38031;
Fax: + 43 1 4277 38034

Abstract

The importance of establishing equivalence in measurement prior to conducting substantive cross-country comparisons has long been emphasized in international business research. However, all approaches currently available for the investigation of measurement equivalence or invariance (ME/I) assume that the focal measures are reflective in nature. In our paper we offer guidelines for the assessment of ME/I when the focal construct is measured with formative indicators. We first identify distinct types of ME/I that are conceptually consistent with the nature of formative measurement models. We then describe a step-by-step procedure that can be employed to test for these types of ME/I in a comparative context. Lastly, we illustrate the application of the proposed procedure with data from a three-country survey on consumers' product-country image perceptions.

Journal of International Business Studies (2010) 41, 360–370.

doi:10.1057/jibs.2009.37

Keywords: cross-cultural research/measurement issues; equivalency; measurement and scale development issues; structural equation modeling

INTRODUCTION

The need to establish equivalence in measurement prior to undertaking substantive cross-country comparisons has long been a theme in the international business literature (e.g., Bhalla & Lin, 1987; Cavusgil & Das, 1997; Hui & Triandis, 1985; Schaffer & Riordan, 2003). Failure to establish measurement equivalence or invariance (ME/I) "is a potential source of measurement error (i.e., discrepancies of what is intended to be measured and what is actually measured), which accentuates the precision of estimators, reduces the power of statistical tests of hypotheses, and provides misleading results" (Hult et al., 2008: 1028).¹

Over the years, several approaches have been proposed to assess ME/I, ranging from relatively simple methods such as profile analysis (e.g., Morris & Pavett, 1992) to more sophisticated procedures such as item response theory (IRT) (e.g., De Jong, Steenkamp, & Fox, 2007; Meade & Lautenschlager, 2004; Raju, Laffitte, & Byrne, 2002) and, most commonly, multi-group confirmatory factor analysis (CFA) (e.g., Mullen, 1995; Myers, Calantone, Page, & Taylor, 2000; Singh, 1995; Steenkamp & Baumgartner, 1998). While such approaches have been put to

Received: 12 February 2008

Revised: 16 January 2009

Accepted: 3 March 2009

Online publication date: 27 August 2009

good use by international business researchers seeking to assess the ME/I of their measures, they are suitable only for measures composed of *reflective* (i.e., effect) indicators. Such measures conform to the domain-sampling model of classical test theory (see Nunnally & Bernstein, 1994), whereby indicators are seen as being effects of the underlying latent variable such that “a shift in the construct leads to an expected shift in an indicator” (Bollen & Ting, 2000: 3).

However, not all constructs are best operationalized by means of reflective indicators, because “in many cases, indicators should be seen as causing rather than being caused by the latent variable measured by the indicators” (MacCallum & Browne, 1993: 533). Such indicators are known as *formative* (or causal), whereby “the direction of causality flows from the indicators to the latent construct, and the indicators, as a group, jointly determine the conceptual and empirical meaning of the construct” (Jarvis, MacKenzie, & Podsakoff, 2003: 201). In the international business literature formative indicators have been used to operationalize such constructs as export performance (Diamantopoulos, 1999), environmental pressures (Venaik, Midgley, & Devinney, 2004, 2005), industry drivers (Johansson & Yip, 1994), export coordination (Diamantopoulos & Siguaw, 2006), and psychic distance (Brewer, 2007); indeed, Venaik et al. (2005: 670) explicitly state that “we would recommend that international business researchers consider the formative model – in a field that studies such broad, diverse and complex phenomena as the global business environment it may make more sense.”

Choosing the correct measurement model (i.e., reflective vs formative) is very important, as measurement model mis-specification results not only in misrepresentation of the epistemic relationships between constructs and their measures (MacKenzie, 2003) but also in biased parameter estimates of structural relationships (Jarvis et al., 2003; Law & Wong, 1999; MacKenzie, Podsakoff, & Jarvis, 2005). Note, in this context, that formative indicators have very different properties from reflective indicators. As a result, “conventional procedures used to assess the validity and reliability of scales composed of reflective indicators (e.g., factor analysis and assessment of internal consistency) are not appropriate for composite variables (i.e., indexes) with formative indicators” (Diamantopoulos & Winklhofer, 2001).² However, it is precisely on extensions of such procedures (e.g., multi-group CFA) that current approaches

for assessing cross-national ME/I are based. For example, in their comprehensive review of methodological contributions to evaluating ME/I, Vandenberg and Lance explicitly state that “the concern is with effect indicators and not causal indicators” (2000: 10). Similarly, approaches based on IRT also invoke reflective measurement assumptions since “the IRT framework posits a log-linear, rather than a linear, model to describe the relationship between observed item responses and the level of the *underlying* latent trait ... CFA and IRT methods of testing for ME/I are conceptually similar but distinct in practice” (Meade & Lautenschlager, 2004: 363, 367; emphasis added).

In short, the current methodological arsenal for assessing ME/I does not cover formative measures. However, formative measures are frequently used to operationalize latent variables in structural equations models (SEM) (Diamantopoulos, Riefler, & Roth, 2008; Williams, Edwards, & Vandenberg, 2003), and assessing their ME/I is important, because “cross-national differences in relationships between scale scores could indicate real differences in structural relations between constructs, or scaling artifacts, ..., or even nonequivalence of the constructs involved” (Steenkamp & Baumgartner, 1998: 78). Construct inequivalence, in particular, poses a very real danger in comparative studies employing formative measures, because “the full meaning of the composite latent construct is derived from its measures” (MacKenzie et al., 2005: 711). Consequently, when the focus is “on relationships among variables and attempts to identify similarities and differences in these relationships across cultures” (Van de Vijver & Leung, 1997: 21), it is essential to ensure ME/I of the relevant formative measures.

Against this background, the purpose of this study is to provide international business researchers with guidelines on the assessment of ME/I in situations where the focal construct is measured with formative indicators. We first highlight some conceptual and statistical issues, and then present a systematic procedure for assessing ME/I that can be applied whenever data are obtained from at least two countries³ and a set of formative indicators is used to operationalize the construct of interest. Next, we illustrate the application of the proposed procedure on data drawn from a cross-national survey of consumers, and we conclude the paper with some suggestions for future research.

FORMATIVE MEASUREMENT AND INVARIANCE

Conceptual Issues

Assuming that m countries are to be compared, the relationship between a composite latent variable η and a set of formative indicators x_i ($i=1, 2, \dots, n$) can be described by the following measurement model:

$$\eta_k = \gamma_{k1}x_1 + \gamma_{k2}x_2 + \dots + \gamma_{kn}x_n + \zeta_k \quad (1)$$

where γ_{ki} is a parameter indicating the contribution of x_i to η in country k ($k=1, 2, \dots, m$), and ζ_k is a disturbance term representing all other causes of η_k not included in the model. It is assumed that $\text{COV}(x_{ki}, \zeta_k)=0$ for all i and that $E(\zeta_k)=0$. Thus the formative measurement model is conceptually very similar to a multiple regression model with n predictors.

Inspection of Eq. (1) reveals that none of the types of cross-national ME/I commonly discussed in connection with reflective scales are applicable when formative indicators are involved.⁴ For example, factorial or configural invariance requires that “the items comprising the measurement instrument should exhibit the same configuration of salient and non-salient factor loadings across different countries” (Steenkamp & Baumgartner, 1998: 80). However, as can be seen from Eq. (1), there is no factor structure specified in a formative measurement model. Other types of ME/I typically invoked within a reflective measurement framework (e.g., invariant factor variances and/or covariances) are equally inappropriate in a formative measurement context. The reason why existing approaches to assessing ME/I are not applicable is that formative indicators are *exogenous* variables (MacCallum & Browne, 1993), and have no individual parameters (such as item loadings and error variances) associated with them (Diamantopoulos, 2006; Edwards & Bagozzi, 2000). The only parameter for each indicator is its γ -coefficient *within* the overall formative model (see Eq. (1)).

In light of the above, we propose the following three types of ME/I for formative measures:

1. *Structure invariance*. This refers to the extent to which the same pattern of salient (non-zero) γ -coefficients defines the structure of the formative measure in different countries. However, the magnitude of these coefficients can vary across countries, as can the residual variance, $\psi_k=\text{VAR}(\zeta_k)$. This is the weakest form of invariance, in that no cross-country equality constraints are imposed on the parameters of the

measurement model in Eq. (1). Given that, in formative measurement, the meaning of the composite latent variable is dependent on its indicators, the absence of structure invariance essentially means that the very nature of the construct is different across countries. Obviously, in this case, it makes no sense to undertake any cross-national comparisons of structural relationships or conduct further (stricter) tests of invariance such as those discussed below. The data have to be analyzed separately on a country-by-country basis.

2. *Slope invariance*. This refers to the extent to which, for each indicator x_i , the magnitude of the associated coefficient, γ_{ki} , is the same across different countries (i.e., $\gamma_{1i}=\gamma_{2i}=\dots=\gamma_{mi}$, for all i). Slope invariance implies that the composite latent variable is influenced to the same extent by a given indicator in each country. If slope invariance is supported, then it is legitimate to compare structural relationships involving the focal construct across countries. If, on the other hand, slope invariance cannot be supported, this indicates that the theoretical importance⁵ of the indicators is not stable across countries. For example, indicator x_1 may have a stronger influence on η in country A than in country B, whereas for, say, indicator x_3 , the reverse might be the case.
3. *Residual invariance*. This refers to the extent to which the variance of the disturbance term, ζ_k , is the same across countries (i.e., $\psi_1=\psi_2=\dots=\psi_m$). Residual invariance implies that the same amount of the variance in the composite latent variable is jointly accounted for by the n indicators across the m different countries. In principle, residual invariance can be assessed independently of slope invariance. However, there is little point in conducting a residual invariance test if slope invariance cannot be established. Specifically, if the individual indicators do not function equivalently across countries, whether the same amount of the composite latent variable's variance is accounted for by them as a group is unlikely to be of interest. On the other hand, evidence of both slope *and* residual invariance can be interpreted as indicating that the formative measure functions identically across countries.

Note that “in practical applications, full measurement invariance frequently does not hold, and the researcher should then ascertain whether there is at least partial measurement invariance” (Steenkamp

& Baumgartner, 1998: 81). The notion of partial invariance is particularly relevant when slope invariance is assessed, since it is unlikely that each indicator in a formative measure will have coefficients of the same magnitude in *all* countries, or that this will apply to *all* indicators (particularly when the number of indicators and/or the number of countries involved is large). When some (but not all) of the γ -coefficients turn out to be invariant across countries, one can speak of *partial slope invariance*. Clearly, the more γ -coefficients that are found to be invariant, the stronger the evidence that the entire formative measure is reasonably invariant.

Statistical Issues

While the types of ME/I discussed above are conceptually consistent with the nature of formative measurement, their empirical assessment is complicated by the fact that, on its own, the formative measurement model is statistically underidentified (Bollen & Lennox, 1991). As a result, none of the country-specific models in Eq. (1) can be estimated and subsequently compared in terms of structure, slope, and residual invariance.

To solve the identification problem, *additional* information needs to be introduced to enable the estimation of the γ -coefficients and the variance of the residual term, ψ_k . While there are several options for achieving identification of formative models (for a review see Diamantopoulos et al., 2008), we concur with Jarvis et al. that “the best option for resolving the identification problem ... is to add two reflective indicators to the formative construct, when conceptually appropriate” (2003: 213). By doing this, the model essentially becomes a multiple indicators – multiple causes (MIMIC) model (Jöreskog & Goldberger, 1975), which is identified. Its formal specification, for each country k , consists of Eq. (1) discussed earlier *plus* the following two equations:

$$y_1 = \lambda_{k1}\eta_k + \varepsilon_{k1} \quad (2)$$

and

$$y_2 = \lambda_{k2}\eta_k + \varepsilon_{k2} \quad (3)$$

Equations (2) and (3) indicate that y_1 and y_2 are congeneric measures of η_k , where λ_{k1} and λ_{k2} are the loadings of y_1 and y_2 on η_k , and ε_{k1} and ε_{k2} represent measurement error (with $\text{COV}(\varepsilon_{k1}, \varepsilon_{k2})=0$ and $E(\varepsilon_{k1})=E(\varepsilon_{k2})=0$).

The choice of y_1 and y_2 needs to be such that they constitute content-valid indicators of the focal composite latent variable η , so that the latter can indeed be viewed as “as a single latent construct with a mixture of formative and reflective indicators rather than as a single reflective-indicator latent construct with multiple causes” (MacKenzie et al., 2005: 727). Moreover, it is essential that y_1 and y_2 are at least metrically invariant across countries,⁶ because the generated estimates of the γ -coefficients and residual variance (ψ_k) are partly dependent on the *specific* external information that is introduced to identify the formative model (Diamantopoulos, 2006; Franke, Preacher, & Rigdon, 2008; Howell, Breivik, & Wilcox, 2008). If the y -variables in the MIMIC model are not invariant, then observed differences in the γ_{ik} and ψ_k estimates cannot be unambiguously attributed to lack of invariance in the formative indicators. By ensuring the metric invariance of the y -indicators *before* testing for structure, slope, and residual invariance, the context-specificity of formative measurement is controlled for.⁷

Bearing the above in mind, we now describe a stepwise procedure designed to help researchers evaluate the three forms of ME/I previously discussed.

A PROCEDURE FOR ASSESSING THE CROSS-NATIONAL INVARIANCE OF FORMATIVE MEASURES

Our proposed procedure involves:

1. testing for metric invariance of the y -variables;
2. estimating a baseline MIMIC model (to establish structure invariance); and
3. introducing equality constraints on the parameters of the x -indicators and the residual term across countries (to test for slope and residual invariance respectively).

Preliminary Tests: Metric Invariance Assessment

The first step in our procedure involves the testing of the metric invariance of the y_1 and y_2 variables to be subsequently used as reflective items in the MIMIC model. To do this, and bearing in mind that a two-indicator single-factor model is not identified (Long, 1983), we specify a *three*-indicator factor model and constrain the loadings of y_1 and y_2 to be equal across the m countries (i.e., we set $\lambda_{11}=\lambda_{21}=\dots=\lambda_{m1}$ and $\lambda_{12}=\lambda_{22}=\dots=\lambda_{m2}$). We use the third item, y_3 , as a reference (marker) indicator to scale the latent variable by setting its loading in all countries equal to 1.⁸



To evaluate the fit of the above model across the m countries, we examine the relevant χ^2 statistic in conjunction with other fit indices recommended in the methodological literature (see, e.g., Hu & Bentler, 1998, 1999; Marsh, Balla, & Hau, 1996): the root mean square error of approximation (RMSEA); the non-normed fit index (NNFI); the comparative fit index (CFI); and the standardized root mean square residual (SRMR).

Assuming that the overall picture painted by the values of these indices supports metric invariance, one can move on to an examination of structure invariance (see next section) and legitimately use γ_1 and γ_2 as reflective items in the MIMIC model. However, if metric invariance is *not* supported, then γ_1 and/or γ_2 has to be replaced with other reflective items that can be shown to be metrically equivalent. Although this can be cumbersome, for reasons already stated, it makes little sense to test for structure, slope, and residual invariance when the γ -variables are themselves not (metrically) invariant.

Baseline Model: Structure Invariance Assessment

The second step in our procedure involves setting up a baseline MIMIC model, whereby no cross-country constraints are imposed on the model parameters, other than on the loadings of γ_1 and γ_2 . The fit of the baseline model can be evaluated by using the various fit indices mentioned previously. Assuming that the fit is reasonable, one can proceed to an investigation of slope invariance; otherwise, no further tests of ME/I are meaningful (since the basic nature of the construct is different in the countries under consideration).

Equality Constraints I: Slope Invariance Assessment

Full slope invariance can be assessed by placing cross-country equality constraints on all γ -parameters of the baseline MIMIC model (i.e., by setting $\gamma_{1i} = \gamma_{2i} = \dots = \gamma_{mi}$ for all i). The resulting MIMIC model is nested within the baseline MIMIC model discussed above, because it can be obtained from the latter by constraining the previously free γ_{ki} parameters to be equal across countries. Therefore a comparison of its fit in relation to that of the baseline model indicates the extent to which the hypothesis of full slope invariance is tenable. Statistically, this can be done by means of the χ^2 difference test (D^2) with degrees of freedom equal to the difference in the degrees of freedom between the two models (Steiger, Shapiro, & Browne, 1985)

and by looking at changes in NNFI, RMSEA, and CFI values.

If the hypothesis of full slope invariance is rejected, then inspection of the modification indices (MIs) and expected parameter changes (EPCs) should help identify which γ -parameters should be freed across countries (Kaplan, 1989). By freeing one γ -parameter at a time and comparing the fit of the less constrained model with those of the full invariance model, *partial slope invariance* can be assessed. Partial slope invariance would be indicated if the fit of the model with some (ideally most) of the γ -parameters set to be equal across countries is (a) significantly better than the fit of the full slope invariance model, and (b) not significantly worse than the fit of the baseline (i.e., fully unconstrained) model.

Equality Constraints II: Residual Invariance Assessment

Residual invariance can be assessed by placing additional cross-country equality constraints on the variances of the disturbance terms of the slope invariance model (i.e., by setting $\psi_1 = \psi_2 = \dots = \psi_m$). Support for it would be indicated by the fit of the model containing equality constraints on both the γ - and the ψ -parameters not being significantly worse than the fit of the slope invariance model as well as that of the baseline model.

APPLICATION OF THE PROPOSED PROCEDURE TO CROSS-NATIONAL SURVEY DATA

Focal Construct

To illustrate our invariance testing procedure, we focus on the product-country image (PCI) construct, which is defined as the "aggregate image for any particular country's products" (Narayana, 1981: 32). This construct was chosen for illustrative purposes for three main reasons: its importance in international marketing research; the frequent use of SEM in PCI studies; and its conceptual appropriateness for formative measurement.

Regarding the first reason, a comprehensive review of research on the PCI theme by Papadopoulos and Heslop (2003) revealed no fewer than 766 studies since the field's inception in the 1960s. Importantly, such research has been published in a wide range of not only international business outlets but also marketing and general management journals. Therefore the PCI construct is relevant to a multitude of researchers and can serve as a good case-in-point for illustrating a

procedure that is of potential interest to many fields of study.

Concerning the use of SEM in PCI research, a recent review (Papadopoulos & Butt, 2005) showed that SEM is used more extensively in PCI studies than in several other areas in international marketing. This review further showed that the frequency of SEM use and the sophistication of the proposed models are increasing over time.⁹

Finally, the conceptual appropriateness of a formative specification for the PCI construct is strongly supported by Roth and Romeo, who note that country images capture the “overall perception consumers *form* of products from a particular country, *based on* their *prior* perceptions of the country’s production and marketing strengths and weaknesses” (1992: 480, emphasis added). Han’s (1989) well known “summary” construct of country image also uses a formative specification whereby consumers’ prior knowledge and experience with brands from a given country form that country’s overall image.

Measures

Based on an extensive review of the relevant literature, we identified a total of 11 distinct attributes relating to a country’s products, which we use as formative indicators of the PCI construct, notably x_1 =workmanship, x_2 =reliability, x_3 =innovativeness, x_4 =quality, x_5 =after-sales service, x_6 =availability, x_7 =appearance, x_8 =value for money, x_9 =technical advancement, x_{10} =known brand names, and x_{11} =product variety. These attributes have been used extensively in previous PCI studies and have been shown to be relevant to respondents, to work together well in different settings and study designs, and to be able to capture the PCI construct domain comprehensively, (see, e.g., Brodowsky, 1998; Heslop, Papadopoulos, Dowdles, Wall, & Compeau, 2004; Laroche, Papadopoulos, Heslop, & Murali, 2005; Nebenzahl & Jaffe, 1996; Papadopoulos, Heslop, & IKON Research Group, 2000). Discussions with several experts on PCI research also confirmed that the above list comprehensively captures the domain of content of the PCI construct, hence supporting content validity.

Each attribute was measured on a seven-point bipolar adjective scale (e.g., good/poor workmanship, reliable/unreliable, good/poor value for money). In addition, three further “global” items were used as reflective indicators of PCI for identification purposes. These were drawn from

Papadopoulos et al. (2000) and captured overall evaluations of PCI (again on a seven-point bipolar adjective format), namely, y_1 =“products from country X are/are not for people like me”, y_2 =“products from country X have/have not what I like to buy”, and y_3 =I would/would not be proud to own products from country X”.

Data Collection and Study Administration

The above measures were incorporated in a questionnaire and administered to samples of consumers in major cities in the UK and Australia, and Hong Kong. Japan was selected as the test origin country because it combines two characteristics that reduce the probability of confounds that might be present in the case of other origins. First, Japanese products are omnipresent, both worldwide and particularly in the selected sample countries. Second, Japanese exports are concentrated in a small number of sectors, such as cars and consumer electronics. Both these characteristics help reduce potential (lack of) familiarity effects on consumer responses.

For the fieldwork, systematic area probability sampling and the “drop-off/pick-up” technique (Brown, 1987) were utilized, following the repeated and successful use of this approach in PCI studies, including many of those cited above and others (e.g., Ahmed & d’Astous, 2003; Balabanis & Diamantopoulos, 2004). The study was coordinated centrally but administered by local colleagues, who participated in the development of the research instrument and pre-tested it at each location to ensure linguistic and conceptual equivalence (Craig & Douglas, 2005; Malhotra, Agarwal, & Peterson, 1996). The fieldwork yielded highly satisfactory response rates of 74% for Australia and Hong Kong and 58% for the UK. Following listwise deletion of missing values, the final sample sizes used for the analysis were 282 for the UK, 286 for Australia, and 325 for Hong Kong.

RESULTS

We first tested for metric equivalence in the y -indicators by specifying a multi-group CFA model, in which y_3 served as the reference (marker) indicator while the loadings of y_1 and y_2 were constrained to be equal across the three countries. The model showed good fit, supporting the hypothesis of metric invariance ($\chi^2(4)=8.628$, $p=0.071$; RMSEA=0.063; NNFI=0.978; CFI=0.990; SRMR=0.045).

Table 1 ME/I test results

Model	χ^2 value	DF	D^2	DF	RMSEA	NNFI	CFI	SRMR
M1: All γ_i 's and ψ 's different	46.516	32			0.039	0.961	0.995	0.0136
M2: All γ_i 's equal	87.532	54	41.016*	22	0.046	0.947	0.988	0.0355
M3: As M2 but γ_4 , γ_7 , and γ_9 free in Hong Kong	60.599	51	14.083	19	0.025	0.983	0.996	0.0185
M4: As M3 plus equal ψ 's	62.589	53	16.073	21	0.025	0.983	0.996	0.0201

* $p < 0.01$; $D^2 = \chi^2$ – difference test in relation to M1 (baseline model).

Table 2 Parameter estimates and standard errors

Indicator	Parameter	Estimated value (Std error)
Workmanship (x_1)	γ_1	−0.023 (0.047)
Reliability (x_2)	γ_2	0.163 (0.034)
Innovativeness (x_3)	γ_3	0.003 ^a (0.023)
Quality (x_4)	γ_4	0.149, (0.051), −0.042 ^a (0.062) (HK)
After-sales service (x_5)	γ_5	0.047 ^a (0.033)
Availability (x_6)	γ_6	0.236 (0.035)
Appearance (x_7)	γ_7	0.145 (0.037), 0.025 ^a (0.044) (HK)
Value for money (x_8)	γ_8	0.195 (0.034)
Technical advancement (x_9)	γ_9	0.071 (0.040), 0.317 (0.057) (HK)
Known brand names (x_{10})	γ_{10}	−0.036 ^a (0.028)
Variety (x_{11})	γ_{11}	0.115 (0.037)
Products from Japan are for people like me (y_1)	λ_1	0.765 (0.054)
Products from Japan have what I like to buy (y_2)	λ_2	1.000 ^b
Residual variance VAR(ζ)	ψ	0.566 (0.080)

^aNot significant (all other parameters significant at $p < 0.05$ or better).

^bFixed parameter.

We next used y_1 and y_2 as reflective indicators in the baseline MIMIC model (M1) to test for structure invariance.¹⁰ As Table 1 shows, the model produced highly acceptable fit statistics, thus supporting structure invariance.

Next, we tested the hypothesis of full slope invariance by constraining all γ -coefficients to be equal across countries. The resulting model (M2) produced a highly significant chi-square value ($\chi^2(54)=87.53$, $p < 0.01$), accompanied by a notable deterioration in all other fit statistics. Moreover, the difference in the chi-square values between the full slope and the baseline model was significant (M2–M1: $D^2(22)=41.02$, $p < 0.01$). Thus the hypothesis of full slope invariance was rejected.

Examination of the MIs and associated EPCs revealed that the significant increase in chi-square between the baseline (M1) and the full slope invariance model (M2) was due to the γ -coefficients of three indicators, namely x_4 (quality), x_7 (appearance), and x_9 (technical advancement); specifically, the EPC statistics indicated that the coefficients of x_4 and x_7 were lower and that of x_9 was higher in Hong Kong than in the other two countries.

Subsequently, we relaxed the equality constraints for the Hong Kong sample in a sequential fashion, starting with the γ -coefficient associated with the highest MI. The resulting partial slope invariance model (M3) was not significantly worse than the baseline model, as indicated by the chi-square difference test (M3–M1: $D^2(19)=14.08$, $p > 0.10$). Moreover, a comparison between the full and partial slope invariance models showed a highly significant improvement in chi-square (M2–M3: $D^2(3)=26.93$, $p < 0.001$). Thus partial slope invariance (with only three invariance constraints relaxed and all in a single country) was supported.

Finally, we also added equality constraints on the variances of the disturbance terms and compared the resulting model (M4) with those of the baseline model (M4–M1: $D^2(21)=16.07$, $p > 0.10$) as well as with the partial invariance model with no constraints on the disturbances (M4–M3: $D^2(2)=1.99$, $p > 0.10$). The results supported the hypothesis of residual invariance under conditions of partial slope invariance.

Table 2 shows the parameter estimates for our final model (M4). As already stated, the coefficients

for only three indicators are not invariant across all three countries, and in all cases the lack of invariance can be attributed to the Hong Kong sample. Specifically, while both x_4 (quality) and x_7 (appearance) are significant attributes forming Japan's PCI in the eyes of UK and Australian consumers, they are apparently of no importance to Hong Kong consumers (the coefficients of x_4 and x_7 are not significant in the Hong Kong sample). Moreover, Hong Kong consumers appear to place greater emphasis on technical advancement (x_9) as compared with the other two samples. Taken together these differences suggest that the PCI construct (at least for Japan as the focal country) is characterized by some variations in its composition in different countries.

DISCUSSION AND CONCLUSIONS

Building on the recent interest in formative measurement, this study adds to the literature on cross-national ME/I by providing researchers with guidelines regarding the assessment of different types of ME/I when the focal construct is operationalized by means of formative indicators. Our contribution is threefold. First, we show that both the types of ME/I associated with reflective measurement and the methods used to test for them are inappropriate when formative indicators are involved. Second, we identify three distinct types of ME/I (structure, slope, and residual invariance) that are conceptually congruent with the nature of the formative measurement model and also highlight statistical issues associated with their assessment. Third, we outline a step-by-step procedure that researchers can follow to test the ME/I of their formative measures and illustrate the application of the procedure on actual survey data from a three-country study.

Contrasting the assessment of ME/I for reflective vs formative measures, several important differences can be noted. First, unlike reflective measurement, in which constructs are assumed to exist independently of their indicators (Borsboom, Mellenbergh, & van Heerden, 2003), in formative measurement models constructs are *specified* as a function of particular indicators chosen to measure them (Edwards & Bagozzi, 2000). The implication of this from an ME/I perspective is that any lack of invariance associated with formative indicators implies that the construct *itself* differs across countries. The degree and substantive importance of this difference depend directly on the source of the invariance. At the one extreme it could reflect a

lack of structure invariance, as evidenced by poor fit of the formative measurement model *even before* cross-country equality constraints are imposed (i.e., the baseline model itself is not stable across countries). At the other extreme it could simply reflect a lack of slope invariance in only one or two formative indicators, with the rest being fully invariant across countries (not unlike the partial invariance situation observed in our illustrative example). In the latter case it becomes a question of judgment whether, taken together, the non-invariant indicators point to a *major* difference in the composition of the focal construct across countries. Specifically, an indicator may turn out to be non-invariant because:

1. its γ -coefficient is significant in some countries but not in others (e.g., see x_7 in Table 2);
2. its γ -coefficient is significant and of consistent sign in all countries investigated, but the absolute magnitude of the coefficient varies substantially across countries (e.g., see x_9 in Table 2); or
3. it displays a significant γ -coefficient in all countries but with inconsistent signs.

Arguably, case (3) points to a more serious difference in the functioning of the indicator across countries than either case (1) or (2), but further methodological research is clearly necessary to throw more light on this issue. Further research is also needed to establish the boundaries of "acceptable" partial slope invariance since, at present, "there is no agreement, ... , on the proportion of the parameters in the model that should be invariant so that the entire measurement model is considered reasonably invariant" (Yoo, 2002: 389).

A second important difference in the assessment of ME/I for reflective vs formative measures arises from the different form of the relationship linking the indicators to the focal construct. Whereas in reflective measurement models each individual indicator is linked to the (underlying) latent variable by means of a *simple* regression equation, in formative models a *multiple* regression specification (see Eq. (1) above) is used to link the set of indicators to the (composite) latent variable (Bollen & Lennox, 1991). The implication of this from an ME/I perspective is that multicollinearity problems may hinder a proper assessment of slope invariance across countries. This is because some formative indicators may display non-significant γ -coefficients in some countries (but not in others), merely because these indicators are strongly correlated with other indicators included in the measurement



model. To the extent that multicollinearity affects the various country datasets in different ways, tests of slope invariance are just as likely to highlight the effects of multicollinearity as to reveal “true” cross-national similarities/differences in the functioning of the indicators involved. Therefore methodological research specifically addressing the issue of multicollinearity in the context of ME/I assessment would be most welcome.¹¹

A third difference between reflective and formative measures, impacting on the assessment of ME/I, is linked to the identification issue. In our illustrative case, the γ_{ki} and ψ_k estimates in Table 2 have been obtained by using y_1 and y_2 as reflective indicators in the MIMIC model specification. Had we used a different set of reflective indicators, a different set of parameter estimates would have been obtained, and therefore the overall picture of invariance painted might also have been different. Thus assessment of ME/I for formative indicators is, in a sense, context-specific, and therefore “the selection of the ‘external’ (i.e., additional) variables necessary for achieving identification is just as crucial in formative measurement models as in the selection of the formative indicators themselves” (Diamantopoulos, 2006: 15). Although this context-specificity has been seen by some as a general weakness of formative measurement (see Howell, Breivik, & Wilcox, 2007; Howell et al., 2008), Bollen argues that it is possible to deal with it by considering alternative outcome variables: “if the coefficients (γ_s) change depending on the outcome, this is evidence of a structurally misspecified model, and we should question the validity of the model and the appropriateness of the latent variable” (2007: 222). Again, this is an issue in need of further research in an ME/I context.

In conclusion, several noteworthy differences distinguish the assessment of ME/I of reflective vs formative measures, and several important issues are in need of further study. Nevertheless, it is hoped that the procedure outlined in this article will provide an initial platform to help international business researchers evaluate the cross-national stability of their formative measures and thus facilitate the application of the latter in comparative research.

NOTES

¹For discussions of the consequences of lack of ME/I in comparative research, see Mullen (1995), Singh (1995), Steenkamp and Baumgartner (1998) and He, Merz, and Alden (2008).

²For a review of formative indicators and their properties, see Bollen and Lennox (1991) and Diamantopoulos and Winklhofer (2001).

³Note that the proposed procedure is applicable whenever any sets of distinct groups are to be compared. While we focus here on the assessment of ME/I from a cross-national comparison viewpoint, the issue of ME/I is relevant whenever a cross-group comparison is involved (for a wide range of applications of ME/I see Vandenberg & Lance, 2000).

⁴For a comprehensive review of types of invariance when reflective measures are involved see Vandenberg and Lance (2000).

⁵Theoretical importance “refers to the change in the criterion variable based on a given change in the predictor variable, which can be measured using the regression coefficient” (Johnson & LeBreton, 2004: 239).

⁶Metric invariance requires that the loadings of y_1 and y_2 be equal across countries (see Steenkamp & Baumgartner, 1998).

⁷An alternative approach would be to still use a MIMIC model but with y_1 and y_2 representing observed outcomes of η rather than reflective indicators of the latter. For example, Diamantopoulos and Siguaw (2006) use two items capturing the effectiveness of knowledge dissemination within the firm as outcomes of their export coordination construct. Yet another approach is to use two (reflectively measured) constructs as outcomes of η . However, this last approach is potentially more problematic, as: (1) each of these outcome constructs would have to be measured by at least three reflective indicators (so that it can be independently identified); (2) the constructs would have to be unrelated to one another (implying absence of structural links or correlated disturbances); and (3) metric invariance of these constructs would have to be established prior to testing for the ME/I of the focal formatively measured construct.

⁸In multi-group research latent variables should *not* be scaled by standardizing their variances. Indeed, “neither the observed nor the latent variables should be standardized. To compare parameters across groups, the variables must be measured in a common metric for all groups” (Jöreskog & Sörbom, 1989: 238).

⁹Virtually all of these models, however, use reflective operationalizations of the PCI construct (see, e.g., Häubl, 1996; Heslop et al., 2004; Laroche et al., 2005).

¹⁰We used y_2 as the reference (marker) indicator by setting its loading to 1 in all three countries.

¹¹In our empirical example multicollinearity did not (fortunately!) pose a problem in any of the three countries investigated, as indicated by variance inflation factors much lower than the typical cut-off threshold of 10 (see, e.g., Kleinbaum,

Kupper, Muller, & Azhar, 1998). However, in other applications of formative measurement multicollinearity has been shown to be a serious problem (see, e.g., Diamantopoulos & Siguaw, 2006).

REFERENCES

- Ahmed, S. A., & d'Astous, A. 2003. Product-country images in the context of NAFTA: A Canada-Mexico study. *Journal of Global Marketing*, 17(1): 23-43.
- Balabanis, G., & Diamantopoulos, A. 2004. Domestic country bias, country-of-origin effects, and consumer ethnocentrism: A multidimensional unfolding approach. *Journal of the Academy of Marketing Science*, 32(1): 80-95.
- Bhalla, G., & Lin, L. Y. S. 1987. Cross-cultural marketing research: A discussion of equivalence issues and measurement strategies. *Psychology and Marketing*, 4(4): 275-285.
- Bollen, K. A. 2007. Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods*, 12(2): 219-228.
- Bollen, K. A., & Lennox, R. 1991. Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2): 305-314.
- Bollen, K. A., & Ting, K. 2000. Tetrad test for causal indicators. *Psychological Methods*, 5(1): 3-22.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. 2003. The theoretical status of latent variables. *Psychological Review*, 110(2): 203-219.
- Brewer, P. A. 2007. Operationalizing psychic distance: A revised approach. *Journal of International Marketing*, 15(1): 44-66.
- Brodowsky, G. H. 1998. The effects of country of design and country of assembly on evaluative beliefs about automobiles and attitudes toward buying them: A comparison between low and high ethnocentric consumers. *Journal of International Consumer Marketing*, 10(3): 85-113.
- Brown, S. 1987. Drop and collect surveys: A neglected technique? *Marketing Intelligence and Planning*, 5(1): 19-23.
- Cavusgil, T. S., & Das, A. 1997. Methodological issues in empirical cross-cultural research: A survey of the management literature and a framework. *Management International Review*, 37(1): 71-96.
- Craig, S. C., & Douglas, S. P. 2005. *International marketing research*, (3rd ed.). Chichester: John Wiley & Sons.
- De Jong, M. G., Steenkamp, J. B., & Fox, J. P. 2007. Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research*, 34(1): 260-278.
- Diamantopoulos, A. 1999. Export performance measurement: Reflective vs formative indicators. *International Marketing Review*, 16(6): 444-457.
- Diamantopoulos, A. 2006. The error term in formative measurement models: Interpretation and modeling implications. *Journal of Modelling in Management*, 1(1): 7-17.
- Diamantopoulos, A., & Siguaw, J. A. 2006. Formative vs reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, 17(4): 263-282.
- Diamantopoulos, A., & Winklhofer, H. W. 2001. Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2): 269-277.
- Diamantopoulos, A., Riefler, P., & Roth, K. P. 2008. Advancing formative measurement models. *Journal of Business Research*, 61(12): 1203-1218.
- Edwards, J. R., & Bagozzi, R. B. 2000. On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2): 155-174.
- Franke, G. R., Preacher, K. J., & Rigdon, E. E. 2008. Proportional structural effects of formative indicators. *Journal of Business Research*, 61(12): 1229-1237.
- Han, M. C. 1989. Country image: Halo or summary construct. *Journal of Marketing Research*, 26(2): 222-229.
- Häubl, G. 1996. A cross-national investigation of the effects of country of origin and brand name on the evaluation of a new car. *International Marketing Review*, 13(5): 76-97.
- He, Y., Merz, M. A., & Alden, D. L. 2008. Diffusion of measurement invariance assessment in cross-national empirical marketing research: Perspectives from the literature and a survey of researchers. *Journal of International Marketing*, 16(2): 64-83.
- Heslop, L. A., Papadopoulos, N., Dowdles, M., Wall, M., & Compeau, D. 2004. Who controls the purse strings: A study of consumers' and retail buyers' reactions in an America's FTA environment. *Journal of Business Research*, 57(10): 1177-1188.
- Howell, R. D., Breivik, E., & Wilcox, J. B. 2007. Reconsidering formative measurement. *Psychological Methods*, 12(2): 205-218.
- Howell, R. D., Breivik, E., & Wilcox, J. B. 2008. Questions about formative measurement. *Journal of Business Research*, 61(12): 1219-1228.
- Hu, L., & Bentler, P. M. 1998. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4): 424-453.
- Hu, L., & Bentler, P. M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria vs new alternatives. *Structural Equation Modeling*, 6(1): 1-55.
- Hui, H. C., & Triandis, H. C. 1985. Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology*, 16(2): 131-152.
- Hult, G. T. M., Ketchen Jr., D. J., Griffith, D. A., Finnegan, C. A., Gonzalez-Padron, T., Harmancioglu, N., Huang, Y., Talay, M. B., & Cavusgil, S. T. 2008. Data equivalence in cross-cultural international business research: Assessment and guidelines. *Journal of International Business Studies*, 39(6): 1027-1044.
- Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30(2): 199-218.
- Johansson, J. K., & Yip, G. S. 1994. Exploiting globalization potential: US and Japanese strategies. *Strategic Management Journal*, 15(8): 579-601.
- Johnson, J. W., & LeBreton, J. M. 2004. History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7(3): 238-257.
- Jöreskog, K. G., & Goldberger, A. S. 1975. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351): 631-639.
- Jöreskog, K. G., & Sörbom, D. 1989. *LISREL 7: A guide to the program and applications*. Chicago: SPSS Inc.
- Kaplan, D. 1989. Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behavioral Research*, 24(3): 285-305.



- Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Azhar, N. 1998. *Applied regression analysis and other multivariate methods*. Pacific Grove, CA: Duxbury Press.
- Laroche, M., Papadopoulos, N., Heslop, L. A., & Mourali, M. 2005. The influence of country image structure on consumer evaluations of foreign products. *International Marketing Review*, 22(1): 96–115.
- Law, K. S., & Wong, C. S. 1999. Multidimensional constructs in structural equation analysis: An illustration using the job perception and job satisfaction constructs. *Journal of Management*, 25(2): 143–160.
- Long, S. J. 1983. *Covariance structure models: An introduction to LISREL*. Beverly Hills, CA: Sage Publications.
- MacCallum, R. C., & Browne, M. W. 1993. The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, 114(3): 533–541.
- MacKenzie, S. B. 2003. The dangers of poor construct conceptualization. *Journal of the Academy of Marketing Science*, 31(3): 323–326.
- MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. 2005. The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, 90(4): 710–730.
- Malhotra, N. K., Agarwal, J., & Peterson, M. 1996. Methodological issues in cross-cultural marketing research: A state-of-the-art review. *International Marketing Review*, 13(5): 7–43.
- Marsh, H. W., Balla, J. R., & Hau, K. T. 1996. An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques*: 315–353. Mahwah, NJ: Lawrence Erlbaum.
- Meade, A. W., & Lautenschlager, G. J. 2004. A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods*, 7(4): 361–388.
- Morris, T., & Pavett, C. M. 1992. Management style and productivity in two cultures. *Journal of International Business Studies*, 23(1): 169–179.
- Mullen, M. R. 1995. Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies*, 26(3): 573–596.
- Myers, M. B., Calantone, R. J., Page Jr., T., & Taylor, C. R. 2000. An application of multiple-group causal models in assessing cross-cultural measurement equivalence. *Journal of International Marketing*, 8(4): 108–121.
- Narayana, C. L. 1981. Aggregate images of American and Japanese products: Implications on international marketing. *Columbia Journal of World Business*, 16(2): 31–35.
- Nebenzahl, I. D., & Jaffe, E. D. 1996. Measuring the joint effect of brand and country image in consumer evaluation of global products. *International Marketing Review*, 13(4): 5–22.
- Nunnally, J. C., & Bernstein, I. H. 1994. *Psychometric Theory*, (3rd ed.) New York: McGraw-Hill.
- Papadopoulos, N., & Butt, I. 2005. The use of structural modeling in international marketing: A review and analysis. In C. Middleton (Ed.), *Managing in turbulent times*. Toronto: Administrative Sciences Association of Canada, International Business Division.
- Papadopoulos, N., & Heslop, L. A. 2003. Country equity and product-country images: State-of-the-art in research and implications. In S. C. Jain (Ed.), *Handbook of research in international marketing*: 402–433. Cheltenham: Edward Elgar.
- Papadopoulos, N., Heslop, L. A., & the IKON Research Group 2000. *A cross-national and longitudinal study of product-country images with a focus on the US and Japan*, Marketing Science Institute Reports, paper 00–106, 67.
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. 2002. Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87(3): 517–529.
- Roth, M. S., & Romeo, J. B. 1992. Matching product category and country image perceptions: A framework for managing country-of-origin effects. *Journal of International Business Studies*, 23(3): 477–497.
- Schaffer, B. S., & Riordan, C. M. 2003. A review of cross-cultural methodologies for organizational research: A best-practices approach. *Organizational Research Methods*, 6(2): 169–215.
- Singh, J. 1995. Measurement issues in cross-national research. *Journal of International Business Studies*, 26(3): 597–619.
- Steenkamp, J. B., & Baumgartner, H. 1998. Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1): 78–90.
- Steiger, J. H., Shapiro, A., & Browne, M. W. 1985. On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 52(3): 253–264.
- Vandenberg, R. J., & Lance, C. 2000. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1): 4–69.
- Van de Vijver, F. J. R., & Leung, K. 1997. *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.
- Venaik, S., Midgley, D. F., & Devinney, T. M. 2004. A new perspective on the integration-responsiveness pressures confronting multinational firms. *Management International Review*, 44(Special Issue 1): 15–48.
- Venaik, S., Midgley, D. F., & Devinney, T. M. 2005. Dual paths to performance: The impact of global pressures on MNC subsidiary conduct and performance. *Journal of International Business Studies*, 36(6): 655–675.
- Williams, L. J., Edwards, J. R., & Vandenberg, R. J. 2003. Recent advances in causal modeling methods for organizational and management research. *Journal of Management*, 29(6): 903–936.
- Yoo, B. 2002. Cross-group comparisons: A cautionary note. *Psychology and Marketing*, 19(4): 357–368.

ABOUT THE AUTHORS

Adamantios Diamantopoulos (PhD, University of Strathclyde) is Chaired Professor of International Marketing at the University of Vienna (adamantios.diamantopoulos@univie.ac.at). His main research interests are in international marketing and quantitative research methodology, and he is the author of well over 100 refereed journal articles in these areas. He serves on the editorial review boards of 12 academic journals.

Nicolas Papadopoulos (PhD, Athens University of Economics and Business) is Chancellor's Professor and Professor of Marketing and International Business at Carleton University (npapadop@carleton.ca). He has over 200 publications, including several books, in his main research areas, which focus on international strategy and buyer responses to it, and include expansion strategy, place images and branding, the role of culture, and international market systems.

Accepted by Donald Siegel, Consulting Editor, 3 March 2009. This paper has been with the authors for two revisions.