Open

INVITED COMMENTARY

Equivalence in international business research: A three-step approach

Angelo M. Solarino[1] and
Peter J. Buckley[2]

[1] *Durham University Business School, University of Durham, Durham DH1 3LB, UK;* [2] *Centre for International Business, Leeds University Business School, University of Leeds, Leeds LS2 9JT, UK*

Correspondence:
AM Solarino, Durham University Business School, University of Durham,
Durham DH1 3LB, UK
e-mail: angelo.solarino@durham.ac.uk

Abstract
A primary research area within the field of international business (IB) is to establish the extent to which concepts, theories, and findings identified in one country are applicable to other contexts and which are unique and cannot be found in other contexts. Researchers in IB acknowledge the importance of the context in their studies, but the practice of assessing equivalence (or invariance) is not widely diffused within the community. We first discuss the components of equivalence (construct, method, and item equivalence), and we offer a three-step approach to address equivalence in the writing and revision of a paper. We aim to help editors, reviewers, and researchers produce more reliable research and navigate the tension between generalizable relationships and context-specific ones, both theoretically and empirically, before performing analysis and hypothesis testing. We then apply equivalence to the construct of firm economic performance as a case study, but the same logic can be applied to other constructs as well.
Journal of International Business Studies (2023) **54**, 550–567.
https://doi.org/10.1057/s41267-022-00562-2

The online version of this article is available Open Access

INTRODUCTION

A primary research area within the field of international business (IB) has been to establish the extent to which concepts, theories, and findings identified in one country are applicable to other contexts (Cuervo-Cazurra 2012; Sekaran, 1983) or what is inherently unique in a context that cannot be found in other contexts (Teagarden, Von Glinow & Mellahi, 2018). Central to the search for standard or generic relationships or uniquely and fully contextualized relationships is equivalence. Neither fully contextualized nor generalizable relationships between constructs can be achieved unless researchers can clearly distinguish differences across research contexts and show that a construct is context-specific or context-invariant.

Researchers are aware that the measures and constructs used in IB research need to be made comparable across countries to make claims about the study findings. Despite this acknowledgement,

the practice of assessing measurement equivalence (or invariance) is not widely diffused within the community (Hult, Ketchen, Griffith, Finnegan, Gonzalez-Padron, Harmancioglu, & Cavusgil, 2008). For example, we found that only about 15% of studies with a multi-country sample in *JIBS* in 2021 and 2020 assessed the equivalence of the constructs and the instruments. In addition, far fewer authors have assessed whether the methodology of the data collection was comparable across the samples. Therefore, although researchers know that the problem exists, the practice of assessing equivalence across countries has not spread broadly.

Failing to assess equivalence is the source of attenuated estimators, which reduce the power of statistical tests of hypotheses and provide misleading results (Davis, Douglas, & Silk, 1981; van de Vijver & Leung, 1997). This has contributed to a plethora of mixed findings in IB research. Frequently, when there are contrasting findings in the literature that do not converge over time, there is a methodological problem (Boyd, Gove, & Hitt, 2005; Ferguson & Ketchen, 1999; Short, Ketchen, & Palmer, 2002).

On the one hand, the increasing availability of large international archival databases of firms has increased the possibilities for academics interested in cross-country comparisons. Prominent examples are Compustat Global, ORBIS, fDi Intelligence, and IBES, as well as large multi-country surveys such as the International Social Survey Program (ISSP), European Values Study (EVS), World Values Survey (WVS), European Union Statistics, and OECD Databank. While the proliferation of these international sources of data makes multi- and cross-country comparisons easier (Boyd, Gove, & Solarino, 2017), it increases the risk that authors will not ascertain the equivalence of these data. The emergence of big data further complicates the picture, as big data are generally collected in automatic ways, possibly under different conditions, making the assessment of equivalence even more important.

With this commentary, we aim to suggest ways in which editors, reviewers, and researchers can produce more reliable research and navigate the tension between generalizable relationships and context-specific ones, both theoretically and empirically, before performing analysis and hypothesis testing. Therefore, we first discuss the components of equivalence (construct, method, and item equivalence), offering a list of key questions that should be addressed in the writing and revision of a paper.

We then apply equivalence to the construct of firm economic performance because it is one of the most common constructs used in IB, but the same logic can be applied to other constructs and measures, such as R&D intensity or entry modes.

EQUIVALENCE: KEY CONCEPTS

Before turning to a more formal definition of equivalence, we highlight the importance of the comparability of constructs and measures as well as the pitfalls involved when comparability is absent. When a researcher samples data from two or more countries, there is a chance that the results of the analysis might mistakenly make the researcher conclude that an effect is stronger in certain countries than in others. Because of the lack of equivalence in the constructs and measures, subsequent studies that try to replicate the findings are likely to arrive at diverging conclusions, giving rise to mixed findings. Overall, the lack of equivalence threatens the generalizability of the findings and our ability to develop cumulative knowledge.

Equivalence can be absent for a wide variety of reasons. Van de Vijver (1998) proposed a conceptual scheme that organizes the sources of non-equivalence into three categories: construct, method, and item bias. Construct bias is the most fundamental source of non-equivalence and recognizes that a construct itself might have different meanings across countries, posing a fundamental threat to international business research. A lack of theoretical validity results in comparing apples and oranges. Certain concepts have meanings that are strongly nation- or culture-specific. Such concepts are called emic concepts, in contrast to universal or etic concepts (Triandis, 1972). For instance, we know that there are several varieties of capitalism (Hall & Soskice, 2001; Witt & Redding, 2014). Differences in the capitalism system result in managers giving different meanings to firm performance and being satisfied with different performance outcomes (Makino & Yiu, 2014). Comparing firm performance between US and China (e.g., Chan, Makino, & Isobe, 2010) might not be informative, as the meaning of performance to managers in the two countries is different.

Method and item bias relate to measurement validity and can be assessed quantitatively (statistically; Meredith & Teresi, 2006; Riordan & Vandenberg, 1994). Method bias results from three sources of bias that can emerge from the sample, the administration, or the instrument used to

collect the data. Samples might be different because they are taken from different populations, or because the subgroups within each population are not equally represented. Differences in sampling procedures could result in differences in the underlying population. These differences then become erroneously interpreted as substantive differences across countries. Different procedures in survey administration or in the instrument across countries can also affect the quality of the data collected and the results of the analysis. In psychology research, there is consensus that the mean of the data collection can affect the distribution of the responses and non-responses, as well as the response style (Billiet, Koch, & Philippens, 2007; Harzing, 2006). In the context of IB, the dimension of the sampling procedures across populations is particularly relevant (Häder and Gabler, 2003; Heeringa & O'Muircheartaigh, 2010).

Item bias refers to anomalies at the item level. Item performances depend on a person's status on a target dimension and the errors that exist around such a dimension. Errors can arise from poor translations of the items (Harkness, Villar, & Edwards, 2010) across groups, or because the item carries different meanings across subgroups. For instance, Makino and Yiu (2014) revealed that not only the means but also the kurtosis of ROA varies systematically across Asian countries. Consequently, unless these differences were appropriately considered, this item would result in biased regression coefficients.

MEASUREMENT EQUIVALENCE: IMPLEMENTATION
To establish equivalence in cross- and multi-country studies, researchers should follow a three-step approach. Table 1 displays the process and shows the key questions journal editors and reviewers should ask, what authors should do to prove equivalence to journal editors, and what actions authors can take if equivalence cannot be confirmed.

Step 1: Construct Equivalence
To obtain construct equivalence and avoid construct bias, it is crucial to draw thorough insights into cross-country similarities and differences in various phenomena. Concepts or constructs used in management research may be conceptually similar but do not perform the same function in all countries. The first question journal editors and

reviewers should ask authors is whether the study adopted an emic or etic standpoint or where it lies on the emic–etic continuum. Were the authors trying to apply a foreign or universal contract to a local phenomenon, or were they aiming to identify or assess how a local construct is generalizable across multiple countries? An example is stock options. Stock options are tools of corporate governance aimed at aligning the interests of senior managers and shareholders and motivating managers to maximize shareholders' value. Most of the applications of stock options in cross-country studies have shown that they worked well in Anglo-Saxon countries but failed to deliver the same results in Europe (Zattoni, 2007). In China, stock options awarded to managers resulted in increasing tunneling rather than minimizing the agency problem (Jiang, Kling, Bo, & Driver, 2017). The application of stock options as an etic construct with the same meaning that this governance tool has in Anglo-Saxon countries is likely to be the source of diverging findings. Applying a construct without considering the context results in an incomplete assessment.

Second, journal editors and reviewers should challenge authors to clarify the relationship between theory and context and contextualize theories and variables (theories in context) or theorize about context (theories of context; Whetten, 2009). Building on the example of stock options, different findings might be rooted in the structure of the firms awarding stock options: whether the firm has, or not, a controlling shareholders, and on the nature of the controlling shareholder (state, family, etc.). In some countries and in the presence of a dominant shareholder, stock options are often used to reward loyalty rather than incentivize performance (Melis, Carta, & Gaia, 2012). In some circumstances, certain constructs can have partial equivalence – similar, but not identical, meanings. Therefore, journal editors and reviewers should ask authors to clarify the extent of the construct equivalence.

To avoid construct bias, authors should obtain insight into cross-country similarities and differences in various phenomena. They could use expert panels and focus groups to identify cross-national differences (Harding, 2013; Johnson, 1998). An understanding of the legal and societal structures of the countries can help direct the areas of inquiry and find similar contracts. When authors cannot empirically prove that the data are comparable across countries, journal editors should ask them to

Table 1 Steps and key questions for assessing measurement equivalence

Step 1: Construct equivalence	→ Step 2: Method equivalence	→ Step 3: Item equivalence
Key questions for editors and reviewers		
•Did the authors use an emic or etic approach?	Have the researchers followed the appropriate sampling procedures across the populations?	Do the items used to measure the construct have the same meaning across the samples?
•Does the construct have the same meaning across the sample population?	Are the subgroups equally/sufficiently represented?	
•What is the degree of equivalence (partial or total)?	Have the data been collected under the same conditions?	
	What is a "representative sample"?	
	Does the construct show comparable properties across countries (similar patterns of convergent and divergent validity)?	
What the authors should do:		
Consult with an expert on the topic to confirm the (degree of) equivalence	Demonstrate that the samples are comparable	Statistically prove item equivalence (CFA, equality of variance test; exponential random graph (p*); Cluster analysis)
Conduct pilot studies to confirm equivalence	Demonstrate that the effect is not driven by a country with a larger number of observations	
	Demonstrate that the variables are constructed following the same procedures	
	In survey research, assess response style, item wording, reverse items, and issues with common method bias	
What if equivalence cannot be achieved? Authors should:		
Look for alternative constructs	Resample the data	Look for alternative measures
Develop a new instrument	Perform subgroup analysis	Conduct separate analysis for each group
Investigate the emic dimensions of the construct (what are the characteristics of this unique culturally embedded construct?	Investigate the source of method bias	Suggest how to refine the construct
Investigate why the construct has different meanings in different settings		

inquire why it is the case. Exploring why variables across countries carry different meanings could be an important contribution and will help building cumulative knowledge. Journal editors should encourage authors to assess which institutional-level antecedent can be used as a predictor of such differences (Cheung & Au, 2005).

Step 2: Method Equivalence
To minimize the risk of method bias and improve methodological equivalence, journal editors and reviewers should ask authors to define as clearly as possible the precise unit(s) of comparison across which they make final contrasts, and then demonstrate that, first, the samples and subsamples from the populations are similar enough to be compared.

Sample populations with different compositions might affect the results. For example, in a cross-country comparison of the role of women in society that includes countries with limited women's rights, having an equal representation of men and women might result in limited data reliability (Usunier, 1998). Some countries are deeply multicultural. For example, India is made up of highly diversified ethnic, religious, and linguistic groups. In contrast, others are explicitly multicultural, such as Switzerland, which emphasizes the defense of local particularisms also in politics and economics. Then there are cases like African countries where for historical reasons the "ethnic" dimension matters more than the "national" one. Editors and reviewers should always ask authors to clarify what

they mean by "representative sample". Researchers often use random samples of listed multinationals (MNEs) in IB research. However, the differences in the composition of the populations of firms listed in different indexes across the world might have affected the results of the analysis. The population of MNEs on stock exchanges varies greatly across countries (e.g., smaller firms in Italy vs. larger firms in Germany). To avoid method bias, researchers must not only compare sampling procedures (similar sample sizes, stratified samples, etc.) but also understand the distribution of firm characteristics in the population to ensure that they are comparing apples to apples. Mandler, Bartsch, and Han (2021) paid particular attention to balancing the respondents to their survey in terms of sociodemographic to ensure that the subsamples captured the same population across countries, and that the differences were not due to one country having more respondents from a sociodemographic group compared to the other.

Sample equivalence is related not only to the subgroup population within a country but also to large imbalances across group sizes. Groups with larger sample sizes have more weight, masking the lack of equivalence between the groups. The imbalance affects the results of regressions and factorial invariance studies and manifests when one group is twice as large as another (Yoon & Lai, 2018).

Moreover, journal editors and reviewers should ask authors to demonstrate that in surveys and interviews, the administration conditions were comparable across informants. For example, the conditions in which a survey is completed can influence the responses (Edwards, 2008). Extensive literature covers survey and interview administration and describes how to assure honest and comparable responses (Aguinis, Villamor, & Ramani, 2021; Solarino & Aguinis, 2021). Finally, in survey research, authors should supply evidence that the survey was not affected by response style, item wording, reverse items, or issues with common method bias (Podsakoff, MacKenzie, & Podsakoff, 2012).

Step 3: Item Equivalence
Finally, assuming that construct and method biases are absent, journal editors and reviewers should ask authors to demonstrate that item bias is absent. Item bias should be addressed statistically. Table 2 presents a synthetic overview of the four methodologies discussed later in the commentary. Journal editors should recommend one of the following

options when authors have not used one, depending on the nature of the study and the number of countries involved. We offer a guided application of the four methodologies in the next section. We briefly introduce them and describe their strengths and weaknesses.

The first approach is multi-group confirmatory factor analysis (MGCFA). MGCFA assesses whether a construct has the same underlying meaning across groups. This methodology is suitable for testing item equivalence across a limited number of groups (two to three, four at the most), as convergence can be an issue, especially for a comparison with more than two groups (Asparouhov & Muthén, 2014). Furthermore, authors should refrain from testing the equivalence using pairwise comparison between groups because it will increase type I errors. In some circumstances, when authors could not even establish partial equivalence, they analyzed the data separately across groups (Hirst, Budhwar, Cooper, West, Long, Chongyuan, & Shipton, 2008). However, this approach is appropriate only if the purpose of the study does not involve comparing groups (Somaraju, Nye, & Olenick, 2021), and the conclusions are specific for each group and not suitable for making comparisons among them.

The second approach relies on item response theory (IRT), which offers a suitable approach for examining the extent of the differential functioning of each item. Researchers should compare not the mean but the standard deviation of the performance measures across countries (Jebb, Morrison, Tay, & Diener, 2020). This method is particularly suitable for single-item indicators. If the theory supports the use of a specific performance measure, and its standard deviation is similar across countries, then the researcher can use that specific performance measure to perform the cross-country comparison. The difference in the means can then be explained using the predictor variable. This approach has the benefit of having no limitations on the number of groups that can be tested concurrently. However, it can test only one indicator at a time.

Finally, we propose that researchers can adopt two network approaches to test for item equivalence. The first is comparable to a nomological network, where the behavior of the correlations between performance measures should be similar enough to allow for a comparison. The nomological network has been criticized for not being able to provide a practical and usable methodology for

Table 2 Comparison of methodologies

Methodology	Suitable for	Not suitable	Strength	Weakness
CFA	Few countries (2–3)	Single indicator Formative measures	Multiple indicators Latent variables Theory driven	Requires a minimum of 3 indicators for each latent variable
Equality of variance test	Many countries (9+)	Latent variables	Ease of implementation	Can compare a single variable at a time
Exponential random graph (p*)	Few countries (4–8)	Single indicator	Latent variables Can compared an unlimited number of variables	Requires a minimum of indicators (20) for a meaningful network
Cluster analysis	Few countries (4–8)	Single indicator Latent variables	Can compared an unlimited number of variables Can find solutions that other methods cannot	Atheoretical

actually assessing construct equivalence and validity. We propose using exponential random graph (*p**) models as a practical solution. This approach allows for a large number of indicators in the analysis. However, this is also a weakness, as it requires a large number of indicators to assess, build, and simulate the possible correlation networks (at least 20). If the networks created do not support similarity across countries, then researchers should try the second network approach we suggest.

This second network approach consists of using cluster analysis to identify those performance indicators that behave more similarly and thus belong to the same cluster. The benefit of this approach is that it can identify solutions that other methods cannot. The solution cannot be explained theoretically, only empirically. As with MGCFA, these methodologies are not suitable for comparing many groups. However, they can accommodate a larger number of groups than MGCFA. In the next section, we apply construct, method, and item bias to the case of firm performance.

A Practical Example with Organizational Performance

As a practical example, we compared performance measures across all listed firms in Mainland China (hereafter China), Hong Kong-SAR (hereafter Hong Kong), and Singapore from 2009 to 2018. The sample was composed of 2451 firms and 26,145 year observations. We chose this sample because previous studies have often compared Hong Kong and Singaporean firms with Chinese ones

(Carlsson, Nordegren, & Sjöholm, 2005; Eng & Spickett-Jones, 2009; Huang, Kerstein, & Wang, 2018; Song, Zeng, & Zhou, 2021). Moreover, studies using Asian samples have become more numerous over the last decade (Bai, Du, & Solarino, 2018; Boyd & Solarino, 2016).

We chose performance as a case study because it is one of the most commonly assessed variables in business research (Boyd & Solarino, 2016; Combs, Crook, & Shook, 2005). However, the discussion below can be applied to any variable under investigation, whether innovation, job satisfaction, leadership, or something else. We collected data for the most commonly used performance measures, adapting the list from Combs et al. (2005). Table 3 reports the list of performance indicators. Table 4 reports the descriptive statistics of the performance measures by country.

Step 1: Construct equivalence
Previous studies have found that performance is a multidimensional construct (Hamann, Schiemann, Bellora, & Guenther, 2013; Rowe & Morrow, 1999; Tosi et al., 2000). Those studies were single-country studies. Not much has been discussed in the literature about how, and to what extent, the construct of organizational performance is valid and reliable across countries. There are good reasons to expect that organizational performance might not be fully equivalent across countries.

There are several ways for scholars to measure performance. Accounting measures are defined as the historical performance of organizations and are assessed through the use of accounting data

Table 3 Performance indicators

Accounting measures	Growth measures[a]	Market measures
Return on assets*	Sales*	Stock returns*
Return on sales*	Profit*	Market-to-book value (Tobin's Q)*
Return on equity*	Earnings per share (EPS) growth*	Stock price/earnings (P/E)
Operating margin*	EV/EBITDA	EV/EBITDA
Net income*		
EPS*		
Gross margin*		
EBITDA margin*		
EBIT margin*		
79% of all accounting measures*	*77%* of all growth measures*	*75%*of all market measures*
41% overall	*13% overall*	*8% overall*
		62% of total measured*

[a] To smooth the effect of a single year for all growth measures; we used a 3-year average
*Identified by Combs et al. (2005) and included in the study

Table 4 Descriptive statistics for the performance indicators for the period 2009–2018

Variable	China		Hong Kong		Singapore	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Shareholder return	2343.70	11,311.43	1737.91	9504.62	847.71	3176.16
Tobin's Q	1.39	3.59	1.27	3.68	0.92	3.52
PE	71.76	440.45	47.42	468.54	36.28	388.84
Turnover	2,579,044	17,900,000	1,066,636	5,381,329	555,815	2,588,535
EBIT	165,960	1,037,927	127,474	869,109	56,489	203,367
Profits before tax	189,869	1,143,120	158,262	1,025,307	70,356	297,763
Profits after tax	105,379	832,247	114,537	781,819	52,161	221,072
Net income	209,930	1,434,188	113,039	771,140	53,777	241,674
ROE (before tax)	8.42	46.13	3.52	60.10	4.89	59.05
ROA (before tax)	4.47	11.42	2.47	16.06	3.38	14.26
ROE (net income)	5.41	42.21	0.27	59.40	2.97	52.55
ROA (net income)	3.16	10.29	1.18	14.74	2.26	14.07
ROS	− 0.01	2.44	– 2.14	84.67	− 0.15	9.21
Profit margin	9.73	19.73	7.02	25.86	9.73	24.61
Gross margin	32.38	21.53	40.52	26.68	41.29	26.66
EBITDA margin	14.69	21.31	11.51	24.53	15.93	22.13
EBIT margin	9.17	18.78	6.77	24.51	12.78	24.78
Enterprise value/EBITDA	11.82	15.34	9.16	14.32	10.59	12.77
EPS	0.09	0.41	– 0.41	30.21	0.04	0.16
Turnover growth	8.48	582.17	3.00	189.29	7.01	326.71
Net income growth	3.86	293.59	− 0.22	22.19	− 1.20	60.39
EPS growth	− 0.16	6.17	− 0.46	14.77	− 0.25	2.56
Enterprise value/EBITDA growth	0.21	7.41	0.60	20.52	0.33	3.57

published in annual reports (Fryxell & Barton, 1990). When observed longitudinally, they can be interpreted as growth measures. Growth measures, while logically distinct from accounting measures (Hamann et al., 2013), are often based on the latter. Therefore, they suffer from similar issues, and we treat them jointly.

Each country has its own accounting "standards" or generally accepted accounting principles (GAAP). For instance, some countries ask to account for specific balance sheet items in different ways (e.g., inventory and asset depreciation; Ball, Robin, & Sadka, 2008). The adoption of the International Accounting Standards has not improved

the situation, with significant differences remaining across countries (Barth, Landsman, & Lang, 2008; De George, Li, & Shivakumar, 2016). Furthermore, differences arise among managers' use of earnings management techniques (Han, Kang, Salter, & Yoo, 2010) to smooth or accentuate profits in a given year, use which is susceptible to the cultural values that dominate the country. For instance, individualism is positively related to the magnitude of earnings discretion, while uncertainty avoidance is negatively related. Other studies have found that the quality of information reporting in financial statements is not consistent across countries because of differences in how institutions monitor and enforce adherence to reporting standards. In countries where regulatory enforcement is weaker, and penalties for false reporting are minimal (Holthausen, 2009), accounts are more easily manipulated. If the cultural dimension is not considered, accounting measures are only partially equivalent.

Market measures are computed using capital market indicators, such as total shareholder return (TSR) or stock price changes. Stock market performance reflects future opportunities and cash flows, in contrast with accounting returns, which entail a historical perspective. Also market measures differ across countries. Market liquidity, market size, market regulations, and transparency vary across stock markets. Market liquidity and size affect the valuation of stock prices, as shares listed in smaller and less liquid markets have lower valuations (Bleck & Liu, 2007). Moreover, market regulations affect the content of reported financial information (Alford, Jones, Leftwich, & Zmijewski, 1993). Market size, liquidity, and transparency of information are necessary for the correct functioning of the market. In a transparent market, shareholders are able to distinguish "good" from "bad" projects and thus achieve the first-best outcome by liquidating poor projects. Overall, as markets are highly sensitive to local dynamics, especially smaller ones, their indicators are not fully comparable. Finally, hybrid indicators, such as price-earing (PE), Tobin's Q, and the market-to-book ratio, present the problem of both types of accounting and market measures of performance. Overall, we should not expect full equivalence of performance indicators across countries.

Step 2: Method equivalence
The second step consisted of demonstrating that the methodology for collecting the data is the same across countries. This implies that the samples are not particular or skewed toward a particular dimension but are representative of the population. In our data collection, we did not impose any boundary conditions on the sample (industry, number of employees, etc.). We collected data from all listed firms in the Hong Kong, Singapore, and Shanghai stock exchanges' main listings. In our case, we did not collect a representative sample but population data. The Chinese and Hong Kong samples represent 45% and 38% of the overall observation, respectively, while the Singaporean sample represents the remaining 17%. Given the unbalanced samples in terms of observations, the largest samples in the dataset will dominate the "regression outputs" obscuring the contribution of the smaller sample, and the different composition of the samples because the lack of boundary conditions will further bias the analysis and make it harder to replicate the results.

Step 3: Item equivalence
Finally, item bias can be mitigated by assessing which measures behave similarly across countries, as we show in the examples that follow. Measurement equivalence is a property of the instrument used to measure the desired variable and implies that the same concept is measured in the same way across subgroups. In our case, it is the performance of firms across countries. Put differently, this occurs when firms with the same standing on the latent trait (performance) but sampled from different groups (countries) have equal expected observed scores on the assessment (Drasgow, 1987; Mellenbergh, 1989).

MGCFA
The first method we used to test measurement equivalence was MGCFA (Joreskog, 1971; Steenkamp & Baumgartner, 1998). We followed Vandenberg and Lance's (2000) multi-step approach to assess measurement equivalence (or invariance) across groups. The steps include testing for (1) configurational equivalence, (2) metric equivalence, (3) scalar equivalence, and (4) measurement equivalence (or uniqueness equivalence). The first model tested for configurational equivalence with all the performance indicators load on a single factor. The model did not converge. We then tested the model using three latent factors: one for accounting measures, one for market measures, and one for growth measures. Similarly, this model did not converge. As the final step, we followed Cheung and Lau's (2012) recommendation and

Table 5 MGCFA test for item equivalence among performance measures in China, Hong Kong, and Singapore

Model	Configurational equivalence	Metric equivalence
One single factor	No	–
Three factors	No	–
Submodels		–
ROE (before tax); ROA (before tax); ROE (net income); ROA (net income)	No	–
Profit margin; Gross margin; EBITDA margin; EBIT margin	Yes	No
Enterprise value/EBITDA; EPS; Tobin's Q; PE; Shareholder return	No	–
EPS; Tobin's Q; PE; Shareholder return	No	–
EPS; PE; Shareholder return	No	–
Turnover growth; Net income growth; EPS growth	No	–
Turnover growth; Net income growth; Enterprise value/EBITDA growth	No	–

tested a subset of the items. Table 5 reports the different models tested for configurational and metric equivalence. Despite several iterations, we were unable to determine which performance indicators were equivalent across China, Hong Kong, and Singapore.

Equality of variance. The second approach tested for the equality (homogeneity) of the performance indicators' variance across countries. We applied Levene's test. This test is an alternative to the Bartlett's test, which is less sensitive to the skewness of the data (departures from normality). We also included the variations on the test, as suggested by Brown and Forsythe (1974). The results are presented in Table 6. Only a few performance measures had comparable variances across China, Hong Kong, and Singapore: net income growth, turnover growth, and EPS (partially) and EPS growth. The most common performance indicators (e.g., ROA, ROE, Tobin's Q) did not have equal variance across countries. Therefore, only a few measures are equivalent.

A network perspective

The last methods we suggest are rooted in network analysis. The first method consists of building a network across the performance indicators for each country and then testing their similarity using an exponential random graph (p*) approach. Faust and Skvoretz (2002) developed a statistical approach for comparing networks independently from the underlying structure, node characteristics, or other characteristics. This approach is based on a set of parameter estimates to predict tie probabilities and then compare and contrast those statistics across n networks. This approach allowed us to compare structural network effects, escaping the assumption of dyadic interdependence commonly adopted in network analysis. The p* methodology can assess the statistical likelihood of specific network configurations that explicitly model non-independence among dyads by including parameters for structural features that capture hypothesized dependencies among ties as a tool for addressing divergence among networks. In the p* framework, the probability of a digraph G is expressed as a log-linear function of a vector of parameters ϑ,[1] an associated vector of digraph statistics x(G), and a normalizing constant $Z(\vartheta)$:

$$P(G) = \frac{\exp(\vartheta' \cdot x(G))}{Z(\vartheta)}.$$

P* models have several benefits compared to traditional network analysis. They allow for the comparison of networks starting from their parameters (and do not assume that observations are independent), and they accommodate attributes and structural estimates as predictors of a given network (Snijders, Pattison, Robins, & Handcock, 2006). Using this logic, two networks, A and B, are similar if their structural tendencies and degrees are similar. If such a condition exists, it should be possible to predict tie probabilities in one network not only from its own parameter estimates but also from those estimated from the other network. Should the two networks require different ϑ for their estimation, then the networks are different.

In the case study, to assess whether the performance metrics were comparable across countries, we adopted a two-step process. The first step consisted of computing the correlation matrix across the performance indicators in each country to create the "performance network." We selected only the correlations significant at 5% or higher to establish a dyadic relationship between two

Table 6 Levene's test for equal variance among performance measures in China, Hong Kong, and Singapore

	Levene's robust test statistic	Levene's robust test statistic (median)	Levene's robust test statistic (10% trimmed mean)	Suitable for cross-country comparison
Shareholder return	27.49	14.22	14.22	No
	(< 0.001)	(< 0.001)	(< 0.001)	
Tobin's Q	14.80	12.02	12.73	No
	(< 0.001)	(< 0.001)	(< 0.001)	
PE	9.89	5.59	5.82	No
	(< 0.001)	(< 0.003)	(< 0.003)	
Turnover	80.02	32.62	33.81	No
	(< 0.001)	(< 0.001)	(< 0.001)	
EBIT	40.61	20.42	21.00	No
	(< 0.001)	(< 0.001)	(< 0.001)	
Profits before tax	41.07	19.19	19.68	No
	(< 0.001)	(< 0.001)	(< 0.001)	
Profits after tax	28.20	15.68	15.68	No
	(< 0.001)	(< 0.001)	(< 0.001)	
Net income	64.51	25.07	25.48	No
	(< 0.001)	(< 0.001)	(< 0.001)	
ROE (before tax)	44.38	39.41	39.26	No
	(< 0.001)	(< 0.001)	(< 0.001)	
ROA (before tax)	149.53	136.22	135.37	No
	(< 0.001)	(< 0.001)	(< 0.001)	
ROE (net income)	55.39	44.68	44.52	No
	(< 0.001)	(< 0.001)	(< 0.001)	
ROA (net income)	161.31	140.98	139.19	No
	(< 0.001)	(< 0.001)	(< 0.001)	
ROS	25.39	10.06	10.79	No
	(< 0.001)	(< 0.001)	(< 0.001)	
Profit margin	91.57	85.07	89.19	No
	(< 0.001)	(< 0.001)	(< 0.001)	
Gross margin	255.49	192.33	241.62	No
	(< 0.001)	(< 0.001)	(< 0.001)	
EBITDA margin	20.10	18.35	19.56	No
	(< 0.001)	(< 0.001)	(< 0.001)	
EBIT margin	124.67	103.62	111.29	No
	(< 0.001)	(< 0.001)	(< 0.001)	
Enterprise value/ EBITDA	17.78	10.57	10.80	No
	(< 0.001)	(< 0.001)	(< 0.001)	
EPS	3.13	0.95	0.95	**Partial**
	(0.043)	(0.385)	(0.386)	
Turnover growth	0.99	0.25	0.25	**Yes**
	(0.370)	(0.775)	(0.775)	
Net income growth	1.43	0.34	0.35	**Yes**
	(0.239)	(0.708)	(0.707)	
EPS growth	1.01	0.60	0.63	**Yes**
	(0.363)	(0.551)	(0.534)	
Enterprise value/EBITDA growth	4.63	3.47	3.49	No
	(0.01)	(0.031)	(0.030)	

Exact p values are reported in brackets
In bold, non-significant Levene's test for equal variance

performance measures. The performance indicators are the nodes of our network, and the 5% significant relationship is the edge. In the second step, we compared these networks. We did not consider whether the relationship was positive or negative, only if it existed. We used Pnet for the analysis (Wang, Robins, Pattison, & Koskinen, 2009).

Table 7 pP^* models for tie equivalence among performance measures in China, Hong Kong, and Singapore (China as the base country)

Observed	China	Hong Kong	Chinese firms listed in Hong Kong	Singapore	For completeness: Taiwan
Edge	61	26	37	31	39
2-star	475	99	183	137	217
3-star	1171	134	294	217	384
4-star	2021	150	331	271	452
5-star	2546	133	265	255	365
Triangles	102	20	36	22	42
T-ratio					
Edge	**0.069**	− 4.533	− 4.643	− 2.081	0.352
2-star	0.259	− 3.3	− 3.542	− 2.075	0.792
3-star	0.423	− 2.61	− 2.797	− 2.070	1.241
4-star	0.556	− 2.165	− 2.265	− 2.064	1.696
5-star	0.634	− 1.823	− 1.861	− 2.061	2.148
Triangles	0.069	− 2.153	− 2.291	− 2.070	1.731

Networks are similar if the *t*-ratio is smaller than 0.1.

We used the networks generated from the performance indicators to predict the Chinese network of performance metrics (baseline network). The results are presented in Table 7. Only the Chinese network can predict itself. Therefore, the relationships between the performance measures in China could not be predicted by those of firms listed in Singapore or Hong Kong or by the performances of Chinese firms listed on the Hong Kong market. As a robustness check, we also tested whether the Chinese data network could be predicted using data from Taiwan, without success. Overall, this methodology suggests that the performance metrics between China, Hong Kong, and Singapore are not equivalent.

The second approach for identifying among performance measures those that behave more similarly across countries was based, similarly, on creating networks of performance measures, but then identifying which "cliques of performance metrics" appear similar across all countries. We used Ucinet 6 for the analysis, and the results of the clique analysis are displayed in Table 8. The results show that shareholder return, PE, and EBITDA margin belonged to the same clique in all three countries.

This allowed us to provide a solid narrative of why these three variables should be chosen as performance metrics. However, we have no theoretical justification for why these measures are more comparable than the other measures.

DISCUSSION

If the goal of IB research is to examine the extent to which theories, models, and constructs are valid and applicable across countries or cultural contexts, scholars aiming to perform cross-country and multi-country studies should first assess whether equivalence exists for the desired construct and variables. While STEM disciplines and psychology have been at the forefront of establishing equivalence in multi-group/multi-country studies, IB and business research has been lagging behind. This lag contributes to problems related to the credibility of our findings (Aguinis, Cascio, & Ramani, 2017; Bergh, Sharp, Aguinis, & Li, 2017; Byington & Felps, 2017; Rynes, Colbert, & O'Boyle, 2018).

In this article, we raised the issue and suggested a series of steps that journal editors, reviewers, and authors should follow to assess the presence of equivalence and the absence of biases at the construct, method, and item levels. The lack of measurement equivalence has substantial implications for IB researchers. First, it can lead researchers to draw inaccurate or, worse, false conclusions, preventing the accumulation of findings and the solidification of knowledge. Many debates in IB, and in business more in general, would benefit from the assessment of measurement equivalence. Recent review papers in JIBS and other IB outlets have highlighted the prevalence of contrasting findings in our field. For example, a recent review of top management teams in international business

Table 8 Clique analysis for clique equivalence among performance measures in China, Hong Kong, and Singapore

Country	Cliques
China	1. 13 cliques found.
	2. EBIT; Profits before tax; Profits after tax; Net income; ROE (before tax); ROA (before tax); ROE (net income); ROA (net income); Profit margin; Gross margin; EBITDA margin; EBIT margin; Enterprise Value/EBITDA; EPS;
	3. Shareholder return; EBIT Profits before tax; Profits after tax; Net income; ROE (before tax); ROA (before tax); ROE (net income); ROA (net income); Profit margin; Gross margin; EBITDA margin; EBIT margin; EPS;
	4. Profits after tax; ROE (before tax); ROA (before tax); ROE (net income); ROA (net income); Profit margin; Gross margin; EBITDA margin; EBIT margin; EPS; EPS growth;
	5. Profits after tax; Gross margin; EBITDA margin; Net income growth;
	6. Turnover; EBIT; Profits before tax; Profits after tax; Net income; Gross margin; Enterprise value/EBITDA; EPS;
	7. Shareholder return; Turnover; EBIT; Profits before tax; Profits after tax; Net Income; Gross margin; EPS;
	8. Profits after tax; Gross margin; Turnover growth; Net income growth;
	9. Profits after tax; Gross margin; Enterprise value/EBITDA; Enterprise value/EBITDA growth;
	10. Profits after tax; Gross margin; EPS growth; Enterprise value/EBITDA growth;
	11. Tobin's Q EBIT; Profits before tax; Net income; EBITDA margin; EBIT margin; Enterprise value/EBITDA; EPS;
	12. Tobin's Q; turnover; EBIT; Profits before tax; Net income; Enterprise value/EBITDA; EPS;
	13. Tobin's Q pe EBIT margin; Enterprise value/EBITDA; EPS;
	14. Shareholder return; PE; EBITDA margin; EPS
Hong Kong	1. 15 cliques found.
	2. Shareholder return; EBIT; Profits before tax; Profits after tax; Net income; ROE (before tax); ROA (before tax); ROE (net income); ROA (net income); Profit margin Gross margin; EBITDA margin; EBIT margin;
	3. Profits before tax; Profits after tax; Net income; ROE (before tax); ROA (before tax); ROE (net income); ROA (net income); Profit margin; Gross margin; Net income growth;
	4. Profits after tax; ROE (before tax); ROA (before tax); ROE (net income); ROA (net income); Profit margin; Gross margin; EBITDA margin; EBIT margin; Enterprise EBITDA;
	5. Profits after tax; ROA (before tax); ROA (net income); Gross margin; Turnover growth;
	6. Profits after tax; Gross margin; EPS;
	7. Profits before tax; Profits after tax; Net income; Gross margin; Net income growth; EPS growth;
	8. Profits after tax; Gross margin; EPS growth; Enterprise value/EBITDA growth;
	9. Profits after tax; Gross margin; Enterprise value/EBITDA; Enterprise value/EBITDA growth;
	10. Shareholder return; Turnover; EBIT; Profits before tax; Profits after tax; Net income; ROE (before tax); ROA (before tax); ROA (net income); Profit margin; EBITDA margin; EBIT margin;
	11. Turnover; Profits before tax; Profits after tax; Net income; ROE (before tax); ROA (before tax); ROA (net income); Profit margin; Net income growth;
	12. Turnover; Profits before tax; Profits after tax; Net income; Net income growth; EPS growth;
	13. Tobin's Q; ROA (before tax); ROE (net income); ROA (net income); Profit margin; EBITDA margin; EBIT margin; Enterprise value/EBITDA;
	14. Tobin's Q; Turnover; ROA (before tax); ROA (net income); Profit margin; EBITDA margin; EBIT margin;
	15. Tobin's Q; PE; EBITDA margin; Enterprise value/EBITDA;
	16. Shareholder return; PE; EBITDA margin;
Singapore	1. 10 cliques found
	2. Shareholder return; Profits before tax; Profits after tax; Net income; ROE (before tax); ROA (before tax); ROE (net income); ROA (net income); Profit margin; Gross margin; EBITDA margin; EBIT margin; Enterprise value/EBITDA; EPS
	3. Shareholder return; EBIT; Profits before tax; Profits after tax; Net income; ROE (before tax); ROA (before tax); ROE (net income); ROA (net income); Profit margin; Gross margin; EBITDA margin; EBIT margin; EPS
	4. Profits before tax; Profits after tax; Net income; ROE (before tax); ROA (before tax); ROE (net income); ROA (net income); Profit margin; Gross margin; EBITDA margin; EBIT margin; EPS; EPS growth
	5. Shareholder return; Turnover; EBIT; Profits before tax; Profits after tax; Net income; Gross margin; EPS
	6. Profits after tax; ROE (before tax); ROE (net income); Gross margin; EBITDA margin; EPS; Net income growth; EPS growth
	7. Profits after tax; Gross margin; Turnover growth
	8. Profits before tax; ROA (before tax); Gross margin; EBITDA margin; Enterprise Value/EBITDA; Enterprise value/EBITDA growth
	9. Profits before tax; ROA (before tax); Gross margin; EBITDA margin; EPS growth; Enterprise Value/EBITDA growth
	10. Gross margin; Turnover growth; Enterprise value/EBITDA growth
	11. Shareholder return; Pe; EBITDA margin

(Cuypers, Patel, Ertug, Li, & Cuypers, 2022) found that in almost every area of the debate, there are mixed findings. For example, the top management team structure might or might not have a positive relationship with the performance of an international joint venture. Cuypers and colleagues rightfully suggested that we should further investigate the issue. We also add that researchers should also assess whether the lack of convergent findings is due to a lack of equivalence. For example, respondents from international joint ventures often come from different countries and have possibly diverging strategic goals. Therefore, they might not interpret the performance of the international joint venture in the same way. These differences in the responses could be a source of mixed findings due to construct, method, or item bias, rather than due to an underlying latent theoretical issue we have not discovered yet.

Other examples of how the lack of equivalence could be driving the mixed findings in the literature arise from debates around family firms and internationalization. A large amount of attention in explaining differences in findings has been given to family characteristics (Arregle, Chirico, Kano, Kundu, Majocchi, & Schulze, 2021), including resources, compensation practices, and individual-level characteristics of family managers. Much less attention has been given to the equivalence of the measures across studies and multi-country studies, and how the lack of equivalence in the constructs, methods, and items contributes to the confusion on the relationship between family firm and internationalization. For example, studies on entry modes by family firms arrived at diverging conclusions regarding whether family firms entering new markets aim for low-commitment modes to minimize risks (Monreal-Pérez & Sánchez-Marín, 2017; Scholes, Mustafa, & Chen, 2016) or high-commitment modes to maximize control (Abdellatif, Amann, & Jaussaud, 2010; Pongelli, Calabrò, & Basco, 2019). Some researchers have attempted to explore the role of moderators, such as ownership structure (Pongelli, Caroli, & Cucculelli, 2016). A lack of measurement equivalence could be a possible source of these conflicting findings. The contextual factors in host countries that are missed because of the lack of equivalence could make family firms choose one or another type of entry mode.

This lack of measurement equivalence within and between studies has broader implications for replicating findings. Researchers who aim to replicate and build on previous literature first need to replicate the original study and then assess whether the research findings are due to methodological artefacts or are substantive. A lack of equivalence raises the possibility that there are biases in published research comparing outcomes across countries (see "Making AIB and IB Relevant and Legitimate" in *AIB Insights* 17(2)), making replication difficult, if not impossible.

Measurement of Non-equivalence as a Research area

International business research is often presented with the tension between context specificities and attempts to create generalizable knowledge that can be applied to other research contexts. The lack of accurate assessments in equivalence has resulted in mixed and contrasting findings that have impaired the field's ability to generate cumulative knowledge. In contrast, assessing the non-equivalence of constructs across countries could be a fruitful avenue of research for IB scholars. It will help researchers uncover truly emic dimensions and false etic constructs and understand, when theorizing about a phenomenon, what is what is truly etic and what is "contextually specific". Theories could then be developed by considering false etic constructs and exploring their source of non-equivalence. This would help researchers generate studies that capture underlying effects with a greater degree of precision and avoid ecological fallacies by assessing and verifying that the underlying assumptions are comparable across countries and samples.

For example, comparative studies on board independence have failed to arrive at conclusive findings about the extent to which board independence matters to firm performance (Dalton, Daily, Ellstrand, & Johnson, 1998; Mutlu, Van Essen, Peng, Saleh, & Duran, 2018). On one side of the equation, there is board independence, which has different meanings in different countries. Some countries have gone all-out on board independence, requiring boards to be made up mostly of independent board members. Other countries have demanded a few independent board members but granted them special and veto powers. Overall, the construct of board independence (and how it is measured today as a percentage or number of independent board members) fails to capture that the independent board in the US has a different role and a different power compared to the one in Europe (Practical Law, 2022). Even worse, in some countries,

independent boards have more a ceremonial than the substantive role that Western theories assign them. Board independence and other "good governance practices" have been adopted in many countries. However, these practices are often not helpful in achieving the desired outcome (Chen, Li, & Shapiro, 2011). Assessing the etic and emic, or localized, value of board independence could be a way to develop context-specific researchers who can inform managers and policymakers. Therefore, researchers should question the emic and etic values of the construct and develop more nuanced theories that account for its etic or emic value. On the other side of the equation is firm performance. As discussed previously, this construct has limited equivalence across countries. Therefore, it is not surprising that the comparative literature on corporate governance and board independence has yet to find consensus because of the lack of equivalence in the constructs under investigation. To solve this debate, like many others, researchers should identify which elements (of corporate governance) are truly generalizable across countries and which depend on the specific context of a country.

Contribution to the Performance Literature
The discussion of what constitutes "performance" for a firm has puzzled scholars for more than 30 years (Venkatraman & Grant, 1986), and the debate has yet to be settled (Hamann et al., 2013; Richard, Devinney, Yip, & Johnson, 2009). For instance, studies have found that company performance is represented by three, four, or even eight factors (Hamann et al., 2013; Rowe & Morrow, 1999; Tosi et al., 2000) not necessarily strongly related to each other. In IB, measurement of the performance construct across countries is further complicated by the fact that institutions differ between nations, and these differences substantially affect how companies report and disclose performance measures and subsequent market reactions (Kumar & Zattoni, 2016).

We contribute to the literature on the validity of construct performance by highlighting the importance of the context in which performance is assessed. The previously mentioned studies explored the multidimensionality of performance in a single country. Therefore, they were unable to capture how the same construct dimensionality would be transferable to other research settings. We demonstrated empirically that the construct of performance needs to be assessed, keeping in mind that it is partially equivalent across countries, at best, and that researchers need to find which measures are the most suitable for the analysis given the countries under investigation.

Equivalence and Big Data
The issue of equivalence will become more pressing as the use of big data becomes more widespread, as large amounts of data can be collected from people and firms from different countries. Depending on the origin, data processing technologies, and data collection methods, big data present the same issues that we have discussed. Furthermore, due to the automated approach to data collection, these issues are amplified. Big data are developing their own measures of data quality, which include volume, variety, and velocity (Schroeck et al., 2012). IBM defined a fourth dimension of big data quality, *veracity*, which refers to "the level of reliability associated with certain types of data" including "truthfulness, accuracy or precision, correctness" (IBM, 2012; Schroeck et al., 2012). Big data researchers will need to develop an appropriate set of criteria rooted in existing debates on equivalence. Big data collected from authoritarian regimes might not be comparable to those in democratic countries. The conditions under which these data are collected differ, and users behave differently because of the limitations on individual freedoms in authoritarian countries. Existing recommendations suggest the use of automated deception detection techniques to increase objectivity by decreasing potential human bias, credibility tracking tools, and sensitivity to linguistic dimensions (Lukoianova, & Rubin, 2014). Big data researchers in IB will have to bring the use of these tools a step forward by assessing how and whether there are differences in deception, credibility, and sensitivity in expressions in each country.

CONCLUSIONS
Leung and Bond (1989) and Hofstede warned us about the risk of a lack of equivalence, but IB studies have not been fully responsive to such calls. Effective changes in publishing norms will require journal editors to recognize and emphasize that studies are not informative unless the researchers can prove that the variables under investigation carry the same meaning in each country. In many journals, we have seen the use of multi-country samples without an assessment of the extent to which the data collected from these samples were comparable. The problem of the validity of these

multi-country samples is worsened by the use of large datasets and big data, where validity and equivalence issues are rarely discussed. Second, journal editors and reviewers should ask researchers to be more transparent with their methodology. Readers should be able to check whether the data were collected using the same procedures across different countries, and there are no possible sources of bias that could affect the study outcomes. Have the measures been properly calibrated across countries so that they capture the desired property correctly? Are the samples comparable, and therefore, no subsample carried more weight in the analysis, risking possible biased results? Or did the data come from similar sources so that the data were collected in a similar manner in all countries? Assessing and acknowledging partial, full, or lack of equivalence should be a distinctive feature of IB research. An analysis of equivalence should always precede any analysis in a multi- and cross-country study.

Authors can play a substantive role in the equivalence debate. Multi- and cross-country research in IB should serve the purpose of creating unique and novel insights, generating broader concepts, or identifying local constructs or phenomena that shape business, rather than purely comparing what is similar and what is different across countries. In this regard, authors have an advantage over journal editors in making the context come alive and telling stories that would otherwise be unknown. Extreme situations or boundary conditions on existing theories can act as a prompt for discussing why a study is needed, why the contextual differences matter, and why something works in one country but not in another.

A primary research area within the field of international business (IB) is to establish the extent to which concepts, theories, and findings identified in one country are applicable to other contexts and which are unique and cannot be found in other contexts. Researchers in IB acknowledge the importance of the context in their studies, but the practice of assessing equivalence (or invariance) is not widely diffused within the community. We first discuss the components of equivalence (construct, method, and item equivalence), and we offer a three-step approach to addressed equivalence in the writing and revision of a paper. We aim to help editors, reviewers, and researchers produce more reliable research and navigate the tension between generalizable relationships and context-specific ones, both theoretically and empirically, before performing analysis and hypothesis testing. We then apply equivalence to the construct of firm economic performance as a case study, but the same logic can be applied to other variables as well.

OPEN ACCESS

NOTES

[1]Different models can use different profiles of digraph properties.

REFERENCES

Abdellatif, M., Amann, B., & Jaussaud, J. 2010. Family versus nonfamily business: A comparison of international strategies. *Journal of Family Business Strategy,* 1(2): 108–116.

Aguinis, H., Cascio, W. F., & Ramani, R. S. 2017. Science's reproducibility and replicability crisis: International business is not immune. *Journal of International Business Studies,* 48: 653–663.

Aguinis, H., Villamor, I., & Ramani, R. S. 2021. MTurk research: Review and recommendations. *Journal of Management,* 47(4): 823–837.

Alford, A., Jones, J., Leftwich, R., & Zmijewski, M. 1993. The relative informativeness of accounting disclosures in different countries. *Journal of Accounting Research,* 31(Supplement): 183–223.

Arregle, J. L., Chirico, F., Kano, L., Kundu, S. K., Majocchi, A., & Schulze, W. S. 2021. Family firm internationalization: Past research and an agenda for the future. *Journal of International Business Studies,* 52(6): 1159–1198.

Asparouhov, T., & Muthén, B. 2014. Multiple-group factor analysis alignment. *Structural Equation Modeling,* 21: 495–508.

Bai, T., Du, J., & Solarino, A. M. 2018. Performance of foreign subsidiaries "in" and "from" Asia: A review, synthesis and research agenda. *Asia Pacific Journal of Management,* 35(3): 607–638.

Ball, R., Robin, A., & Sadka, G. 2008. Is accounting conservatism due to debt or equity markets? An international test of "contracting" and "value relevance" theories of accounting. *Review of Accounting Studies,* 13: 168–205.

Barth, M. E., Landsman, W. R., & Lang, M. H. 2008. International accounting standards and accounting quality. *Journal of Accounting Research,* 46(3): 467–498.

Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. 2017. Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization,* 15(3): 423–436.

Billiet, J., Koch, A., & Philippens, M. 2007. Understanding and improving response rates. In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (Eds.), *Measuring attitudes cross-nationally: Lessons from the European social survey*: 113–138. London: Sage.

Bleck, A., & Liu, X. 2007. Market transparency and the accounting regime. *Journal of Accounting Research,* 45(2): 229–256.

Boyd, B. K., & Solarino, A. M. 2016. Ownership of corporations: A review, synthesis, and research agenda. *Journal of Management,* 42(5): 1282–1314.

Boyd, B. K., Gove, S., & Hitt, M. 2005. Construct measurement in strategic management research: reality or illusion? *Strategic Management Journal,* 26(3): 239–257.

Boyd, B. K., Gove, S., & Solarino, A. M. 2017. Methodological rigor of corporate governance studies: A review and recommendations for future studies. *Corporate Governance: an International Review,* 25(6): 384–396.

Brown, M. B., & Forsythe, A. B. 1974. Robust tests for the equality of variances. *Journal of the American Statistical Association,* 69(346): 364–367.

Byington, E. K., & Felps, W. 2017. Solutions to the credibility crisis in management science. *Academy of Management Learning & Education,* 16(1): 142–162.

Carlsson, J., Nordegren, A., & Sjöholm, F. 2005. International experience and the performance of Scandinavian firms in China. *International Business Review,* 14(1): 21–40.

Chan, C. M., Makino, S., & Isobe, T. 2010. Does subnational region matter? Foreign affiliate performance in the United States and China. *Strategic Management Journal,* 31(11): 1226–1243.

Chen, V. Z., Li, J., & Shapiro, D. M. 2011. Are OECD-prescribed "good corporate governance practices" really good in an emerging economy? *Asia Pacific Journal of Management,* 28(1): 115–138.

Cheung, G. W., & Lau, R. S. 2012. A direct comparison approach for testing measurement invariance. *Organizational Research Methods,* 15(2): 167–198.

Combs, J. G., Crook, T. R., & Shook, C. L. 2005. The dimensionality of organizational performance and its implications for strategic management research. *Research Methodology in Strategy and Management,* 2: 259–286.

Cuervo-Cazurra, A. 2012. Extending theory by analyzing developing country multinational companies: Solving the Goldilocks debate. *Global Strategy Journal,* 2(3): 153–167.

Cuypers, I. R., Patel, C., Ertug, G., Li, J., & Cuypers, Y. 2022. Top management teams in international business research: A review and suggestions for future research. *Journal of International Business Studies,* 53: 481–515.

Dalton, D. R., Daily, C. M., Ellstrand, A. E., & Johnson, J. L. 1998. Meta-analytic reviews of board composition, leadership structure, and financial performance. *Strategic Management Journal,* 19(3): 269–290.

Davis, H. L., Douglas, S. P., & Silk, A. J. 1981. Measure unreliability: A hidden threat to cross-national marketing research? *Journal of Marketing,* 45(2): 98–109.

De George, E. T., Li, X., & Shivakumar, L. 2016. A review of the IFRS adoption literature. *Review of Accounting Studies,* 21(3): 898–1004.

Drasgow, F. 1987. Study of the measurement bias of two standardized psychological tests. *Journal of Applied Psychology,* 72(1): 19–29.

Edwards, J. R. 2008. To prosper, organizational psychology should … overcome methodological barriers to progress. *Journal of Organizational Behavior,* 29: 469–491.

Eng, T. Y., & Spickett-Jones, J. G. 2009. An investigation of marketing capabilities and upgrading performance of manufacturers in mainland China and Hong Kong. *Journal of World Business,* 44(4): 463–475.

Faust, K., & Skvoretz, J. 2002. Comparing networks across space and time, size and species. *Sociological Methodology,* 32(1): 267–299.

Ferguson, T. D., & Ketchen, D. J., Jr. 1999. Organizational configurations and performance: The role of statistical power in extant research. *Strategic Management Journal,* 20(4): 385–395.

Fryxell, G. E., & Barton, S. L. 1990. Temporal and contextual change in the measurement structure of financial performance: Implications for strategy research. *Journal of Management,* 16(3): 553–569.

Häder, S., & Gabler, S. 2003. Sampling and estimation. In J. Harkness, F. van de Vijver, & P. Mohler (Eds.), *Cross Cultural Survey Methods* New York: Wiley.

Hall, P. A., & Soskice, D. 2001. *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage.* Oxford: Oxford University Press.

Han, S., Kang, T., Salter, S., & Yoo, Y. K. 2010. A cross-country study on the effects of national culture on earnings management. *Journal of International Business Studies,* 41(1): 123–141.

Harding, S. 2013. Rethinking standpoint epistemology: What is "strong objectivity"? In *Feminist epistemologies*: 49–82. Routledge.

Harkness, J. A., Villar, A., & Edwards, B. 2010. Translation, adaptation, and design. In J. A. Harkness, M. E. Braun, B. E. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts*: 115–140. New York: Wiley.

Harzing, A. W. 2006. Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management,* 6(2): 243–266.

Heeringa, S. G., & O'muircheartaigh, C. 2010. Sample design for cross-cultural and cross-national survey programs. Survey methods in multinational, multiregional, and multicultural contexts, 251–267.

Hirst, G., Budhwar, P., Cooper, B. K., West, M., Long, C., Chongyuan, X., & Shipton, H. 2008. Cross-cultural variations in climate for autonomy, stress and organizational productivity relationships: A comparison of Chinese and UK manufacturing organizations. *Journal of International Business Studies,* 39(8): 1343–1358.

Holthausen, R. W. 2009. Accounting standards, financial reporting outcomes, and enforcement. *Journal of Accounting Research,* 47(2): 447–458.

Huang, H. H., Kerstein, J., & Wang, C. 2018. The impact of climate risk on firm performance and financing choices: An international comparison. *Journal of International Business Studies,* 49(5): 633–656.

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

IBM. 2012. Solutions big data. http://www05.ibm.com/fr/events/netezzaDM_2012/ Solutions_Big_Data.pdf. Accessed 01-03-2022.

Jebb, A. T., Morrison, M., Tay, L., & Diener, E. 2020. Subjective well-being around the world: Trends and predictors across the life span. *Psychological Science,* 31(3): 293–305.

Jiang, L., Kling, G., Bo, H., & Driver, C. 2017. Why do firms adopt stock options and who benefits? A natural experiment in China. *Pacific-Basin Finance Journal,* 46: 124–140.

Kumar, P., & Zattoni, A. 2016. Institutional environment and corporate governance. *Corporate Governance: an International Review,* 24(2): 82–84.

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

Lukoianova, T., & Rubin, V. 2014. Veracity roadmap: Is big data objective, truthful and credible? *Advances in Classification Research Online,* 24(1): 4–15.

Makino, S., & Yiu, D. W. 2014. A survey of strategic behaviour and firm performance in Asia. In M. A. Witt, & G. Redding (Eds.), *The Oxford handbook of Asian business systems*: 561–586. Oxford: Oxford University Press.

Mandler, T., Bartsch, F., & Han, C. M. 2021. Brand credibility and marketplace globalization: The role of perceived brand globalness and localness. *Journal of International Business Studies,* 52(8): 1559–1590.

Melis, A., Carta, S., & Gaia, S. 2012. Executive remuneration in blockholder-dominated firms. How do Italian firms use stock options? *Journal of Management & Governance,* 16(3): 511–541.

Mellenbergh, G. J. 1989. Item bias and item response theory. *International Journal of Educational Research,* 13(2): 127–143.

Meredith, W., & Teresi, J. A. 2006. An essay on measurement and factorial invariance. Medical Care, S69–S77.

Monreal-Pérez, J., & Sánchez-Marín, G. 2017. Does transitioning from family to non-family controlled firm influence internationalization? *Journal of Small Business and Enterprise Development,* 24(4): 775–792.

Mutlu, C. C., Van Essen, M., Peng, M. W., Saleh, S. F., & Duran, P. 2018. Corporate governance in China: A meta-analysis. *Journal of Management Studies,* 55(6): 943–979.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. 2012. Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology,* 63: 539–569.

Pongelli, C., Calabrò, A., & Basco, R. 2019. Family firms' international make-or-buy decisions: Captive offshoring, offshore outsourcing, and the role of home region focus. *Journal of Business Research,* 103: 596–606.

Pongelli, C., Caroli, M. G., & Cucculelli, M. 2016. Family business going abroad: The effect of family ownership on foreign market entry mode decisions. *Small Business Economics,* 47(3): 787–801.

Practical Law. 2022. *Corporate Governance and Directors' Duties.* http://thomsonreuters.com. Accessed 01-03-2022.

Richard, P. J., Devinney, T. M., Yip, G. S., & Johnson, G. 2009. Measuring organizational performance: Towards methodological best practice. *Journal of Management,* 35(3): 718–804.

Riordan, C. M., & Vandenberg, R. J. 1994. A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management,* 20(3): 643–671.

Rynes, S. L., Colbert, A. E., & O'Boyle, E. H. 2018. When the "best available evidence" doesn't win: How doubts about science and scientists threaten the future of evidence-based management. *Journal of Management,* 44(8): 2995–3010.

Scholes, L., Mustafa, M., & Chen, S. 2016. Internationalization of small family firms: The influence of family from a socioemotional wealth perspective. *Thunderbird International Business Review,* 58(2): 131–146.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. 2012. Analytics: The real-world use of big data. How innovative enterprises extract value from uncertain data. http://www.stthomas.edu/gradsoftware/files/BigData_RealWorldUse.pdf. Accessed 01-03-2022.

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

Short, J. C., Ketchen, D. J., Jr., & Palmer, T. B. 2002. The role of sampling in strategic management research on performance: A two-study analysis. *Journal of Management,* 28(3): 363–385.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. 2006. New specifications for exponential random graph models. *Sociological Methodology,* 36: 99–153.

Solarino, A. M., & Aguinis, H. 2021. Challenges and best-practice recommendations for designing and conducting interviews with elite informants. *Journal of Management Studies,* 58(3): 649–672.

Somaraju, A. V., Nye, C. D., & Olenick, J. 2021. A review of measurement equivalence in organizational research: What's old, what's new, what's next? *Organizational Research Methods,* 10944281211056524.

Song, S., Zeng, Y., & Zhou, B. 2021. Information asymmetry, cross-listing, and post-M&A performance. *Journal of Business Research,* 122: 447–457.

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

Teagarden, M. B., Von Glinow, M. A., & Mellahi, K. 2018. Contextualizing international business research: Enhancing rigor and relevance. *Journal of World Business,* 53(3): 303–306.

Triandis, H. C. 1972. *The analysis of subjective culture.* Wiley-Interscience.

Usunier, J. C. 1998. *International and cross-cultural management research.* Sage.

van de Vijver, F., & Leung, K. 1997. *Methods and data analysis of comparative research.* Allyn & Bacon.

Vandenberg, R. J., & Lance, C. E. 2000. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods,* 3(1): 4–70.

Venkatraman, N., & Grant, J. H. 1986. Construct measurement in organizational strategy research: A critique and proposal. *Academy of Management Review,* 11(1): 71–87.

Vijver, F. J. v. d. 1998. Towards a theory of bias and equivalence. In J. Harkness (Ed.), Cross-cultural survey equivalence (pp. 41-65).

Wang, P., Robins, G., Pattison, P., & Koskinen, J. 2009. *PNet: Program for the simulation and estimation of exponential random graph models.* Melbourne: The University of Melbourne, Melbourne School of Psychological Sciences.

Whetten, D. A. 2009. An examination of the interface between context and theory applied to the study of Chinese organizations. *Management and Organization Review,* 5(1): 29–56.

Witt, M. A., & Redding, G. (Eds.). 2014. *The Oxford handbook of Asian business systems.* Oxford, England: Oxford University Press.

Yoon, M., & Lai, M. H. 2018. Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal,* 25(2): 201–213.

Zattoni, A. 2007. Stock incentive plans in Europe: Empirical evidence and design implications. *Corporate Ownership and Control,* 4(4): 54–62.

ABOUT THE AUTHORS

Angelo M. Solarino (PhD, City University of Hong Kong) is Associate Professor at the Durham University Business School, UK. His research focuses on questions of international corporate governance and international business, and the related the methodological issues. His research has been published in the *Journal of Management*, *Journal of Management Studies*, *Corporate Governance: An International Review* and *Strategic Management Journal*, among others.

Peter J. Buckley is Professor of International Business, Centre for International Business, University of Leeds, Visiting Professor of International Business, University of Queensland and Visiting Professor at Hong Kong Metropolitan University, President of the Academy of International Business (2002–04), Chair of the European International Business Academy (2009–2012), Appointed Officer of the Order of the British Empire (OBE) 2012, and elected a Fellow of the British Academy (FBA), 2014.