MARKETING
ZFP – Journal of Research and Management
www.marketing-zfp.de

Research Articles

Multilevel Structural Equation Modelling in Marketing and Management Research

By Johannes Hirschmann and Bernhard Swoboda

Multilevel (or mixed linear) modelling (MLM) either simultaneously tests hypotheses at several levels of analysis or controls for confounding effects at one level while testing hypotheses at another level. Advances in multilevel structural equation modelling (MSEM) enable the specification of latent variables, which are more common in marketing research than the manifest variables employed in hierarchical linear modelling (HLM), and open new conceptual possibilities. However, MSEM involves several challenges and is not frequently used. The authors therefore outline key methodological requirements, options, and challenges regarding MSEM and provide a systematic approach for its use. To achieve this goal, multilevel modelling and the advantages of MSEM over HLM are illustrated, followed by a literature review of marketing and management studies, to determine how MSEM and HLM are differentially applied. The requirements, options, and challenges of MSEM are systematically illustrated by elaborating current knowledge in the literature and by presenting an empirical study describing the sampling, measurement, and methodological issues for three model types: cross-level effects, cross-level interactions, and cross-level effects and interactions. Promising directions and major challenges for future research are examined.

Johannes Hirschmann is Research Assistant and Doctoral Student at the Chair of Marketing and Retailing at Trier University, Universitaetsring 15, 54296 Trier, Germany, Phone: +49/651 201 2607, E-Mail: j.hirschmann@uni-trier.de.
* Corresponding author

Bernhard Swoboda is Professor of Marketing and Retailing at Trier University, Universitaetsring 15, 54296 Trier, Germany, Phone: +49/651 201 3050, E-Mail: b.swoboda@uni-trier.de.

1. Introduction

International business research often involves two levels (e.g., customers, decisions, or subsidiaries' performance nested within countries, cultures, or regions), as do strategy research (e.g., how resources and industry structure affect firm performance) and sales research (e.g., successful sales persons nested within sales teams nested within organizations, with three levels). MLM allows one to accurately model lower-level (level 1) effects *and* the surrounding (level 2) context in addition to various interrelations between both levels. Hierarchical data were previously analysed with fixed-parameter linear regression, which does not account for shared variance. Historically, scholars in the fields of education (e.g., Burstein 1980), biology (e.g., Laird and Ware 1982), sociology (e.g., Blalock 1984), and management (e.g., Mossholder and Bedeian 1983) were the first to discuss MLM. However, only after advances in statistical theory did the application of MLM surge. One of the first applications was in educational research (Aitkin and Longford 1986). The basics of MLM are well known (e.g., Hox 2010), and HLM is frequently employed (a form of ordinary least squares (OLS) regression, analysing variance in the outcome variable with predictor variables at different hierarchical levels). However, HLM does not handle latent constructs and is inferior to MSEM. For example, HLM does not account for measurement errors. Therefore, MSEM is increasingly methodologically discussed (e.g., Barile 2016; du Toit and Du Toit 2008; Hox 2013) but seldom applied. A review of current and future applications and the methodological requirements, options, and challenges associated with MSEM would be beneficial.

In marketing research, latent factors are common, and latent-factor multi-group structural equation modelling (SEM) dominates the analysis of data from independent samples (e.g., Jöreskog 1971). Advances in SEM analysing multiple levels have created unrealized potential (Preacher et al. 2016; Preacher et al. 2010). Currently, methodological papers (e.g., Rabe-Hesketh et al. 2012), textbooks (e.g., Heck and Thomas 2015; Kelloway 2015), and a few studies (e.g., Swoboda et al. 2016; Zhou et al. 2010) have addressed MSEM. The papers are technical and may be difficult to understand; the textbooks are extensive in scope and less suitable as a compact reference, and the studies do not describe MSEM in detail. A systematic, nontechnical guide for conducting an MSEM analysis is lacking, though examples demonstrate the relevance of MSEM and of such a guide.

- First, a strong corporate reputation (CR) is known to affect firms' performance, and multinational firms' increasingly manage their CR across nations to, for example, attract local customers. However, the effects of an often centrally managed but locally perceived CR (i.e., consumers' overall evaluation of a firm's responsibility, strength, or quality of offers, Walsh and Beatty 2007) are likely to vary across nations, giving rise to the question of whether country differences reinforce or diminish CR perceptions and effects. Answering this question allows for better CR management, e.g., in countries with similar reinforcing or diminishing institutions. Scholars have provided insights, particularly regarding the role of culture as a moderator of CR effects on behavioural outcomes (e.g., Bartikowski et al. 2011; Falkenreck and Wagner 2010). However, most studies compared few countries in multi-group models. The three to five countries that were analysed differ not only culturally but also with respect to further country institutions, and the extent to which culture explains CR effects remains unclear. Swoboda et al. (2016) demonstrated differences between the results of multi-group models and those of MSEM. Deephouse et al. (2016) applied HLM to analyse the role of culture on CR perception differences across nations, whereas Swoboda and Hirschmann (2017) used MSEM and obtained contradictory results. The contradictory results are maybe obvious because multi-group models disregard shared variances, for example, and HLM is unable to analyse latent constructs.
- Second, MSM is relevant to sales management. A possible research question is whether and how individual coaching frequency relates to sales managers' goal attainment and how this relationship is reinforced or diminished by regional managers' coaching skills. When district managers' coaching skills are assumed to affect the clarity of the sales teams' roles, two distinct models exist on both levels and cannot effectively be analysed without MSEM (e.g., Dahling et al. 2015). Considering how division managers' leadership styles affect the applicability of regional sales managers' coaching introduces a third level to the model. The prevailing existence of latent constructs (e.g., human perceptions) and the need for aggregate data (e.g., team-level data must be aggregates of individual-level observations because the team itself cannot be questioned) make MSEM relevant in such research. Therefore, in similar human resource management research fields, multilevel theorizing and modelling has gained popularity (e.g., Peccei and van de Voorde 2016).
- Third, MSEM can be used for Big Data in retailing research. A question might be whether perceived price promotions affect patronage behaviour over multiple time points and how this effect is moderated by consumers' price sensitivity. Studies have addressed similar issues without MSEM: Venkatesan and Farris (2012) analysed how price promotions affect purchases but are not affected over time, whereas Yi and Yoo (2011) analysed how sales promotions affect consumers' brand attitudes over time by grouping consumers. Conceptually, perceived price promotions and patronage behaviour can be situated at the time level 1 and price sensitivity can be situated at the consumer level 2. MSEM is necessary because of both latent constructs (perceived price promotions and price sensitivity) and the behaviour measure; additionally, it detects the amount of variance in patronage behaviour that is explained by (more time-invariant) individuals' price sensitivity and by the time of the perceived price promotion. Retail studies have often disregarded the hierarchical data structure (e.g., Hoppner and Griffith 2015).

The examples indicate that MSEM avoids erroneous results and allows for an investigation of research questions and models that would not have been possible otherwise. As MSEM is not frequently used, our first research objective is to introduce scholars to MSEM and its advantages over HLM and to provide novel insights into the topics and shortcomings of MSEM/HLM use in extant studies. We contribute to the literature by providing a review of 527 studies in marketing and management research in which HLM and MSEM are employed. Four categories of multilevel hypotheses found in studies published in 22 leading journals over 20 years are discussed and differentiated into five characteristic research fields (general marketing, international marketing, international management, general management, and human resource management).

Our second objective is to address the requirements, options, and challenges of MSEM. We contribute to the literature by discussing the sampling, measurement, and methodological requirements, options, and challenges of MSEM-based studies. In doing so, we provide a nontechnical explanation and a systematic step-by-step procedure for designing and conducting a cross-sectional MSEM study that tests hypotheses across levels.[1] Additionally, the limitations and oversights are addressed. Our empirical example presents the results for three frequently used types of MSEM models: cross-level effects, cross-level interactions, and cross-level effects and interactions. We provide the first systematic illustration and interpretation to help scholars conduct MSEM-based studies.

The remainder of this article proceeds as follows. After an introduction of MSEM (compared with HLM), the

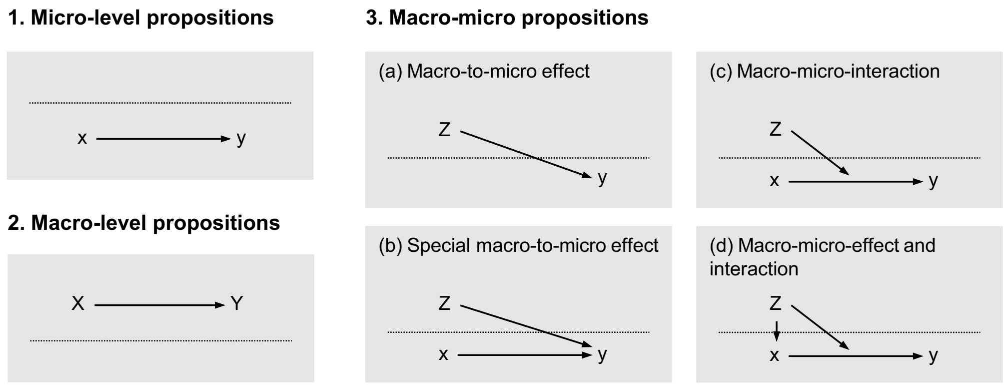

Fig. 1: Categories of possible multilevel propositions

use of both in extant studies is addressed. The methodical requirements, options, and challenges of MSEM are systematically elaborated; sampling, measurement, and methodological issues are addressed; and the empirical results for the three types of MSEM models are presented.

2. Introduction to multilevel structural equation modelling

Understanding MSEM requires knowledge of MLM's basic logic, which we briefly describe before addressing HLM and MSEM. For a hierarchical data structure in MLM, a unit at the lowest (micro) level of measurement must be nested within one unit at a higher (macro) level. Individual-level variables are attitudes, perceptions, or behaviours, for example, which may vary due to economic or political institutions of countries or sizes or leadership styles of firms. MLM is needed if a researcher is interested in propositions that connect variables at different levels (macro/micro) or if a multistage sample design has been employed. MLM has three categories of possible propositions (see *Fig. 1*) (Snijders and Bosker 2012, pp. 10–12).

Micro-level propositions only consider relationships between variables on level 1 (e.g., the individual level). The dotted line indicates that even though all variables of interest are on the micro-level, a macro-level also exists, i.e., the hypothesis does not refer to the macro-level but the macro-level may be present in the sample. For example, one might question how consumers' perceptions of a firm's CR affect trust while sampling consumers from different countries. MLM controls for such differences (e.g., Balka et al. 2014). *Macro-level propositions* consider relationships between variables on level 2. Examining the relationship between the sales team budget and sales team turnover in EUR is not multilevel. In contrast, when either variable or both variables (independent and dependent) are not directly observable on the macro-level and need to be measured at the micro-level and aggregated (e.g., means of individual observations), then a two-stage sample is needed. For example, when examining the relationship between sales team climate and sales team turnover, the sales team climate needs to be aggregated from individual perceptions (Snijders and Bosker 2012). MLM controls for this issue. We do not discuss *micro-level and macro-level views* further.

Macro-micro propositions concern instances in which macro-level variables are related to micro-level variables. Within this category, four model types are common:

(a) *Macro-to-micro effects* (*cross-level effects*), where *Z* has a direct effect on *y*. Referring to the initial example, a researcher might want to model the effect of national culture – a level 2 variable – on consumers' perceptions of a firm's CR – a level 1 variable – to explain CR perception differences across nations. Thus, the effect occurs from one level to another level.

(b) *Special macro-to-micro-effects*, with a relation between *Z* and *y*, where the effect of *x* and *y* is considered. For example, consumers' perceptions of CR (level 1) may simultaneously depend on national culture (level 2) and on consumers' perceptions of a firm's perceived online communication activities (level 1).

(c) *Macro-micro-interaction* (*cross-level interactions*), where the relation between *x* and *y* is dependent on *Z*. Interactions can be included in a multilevel model and can fall between any pair of variables, regardless of the level of conceptualization. For example, the relation between individuals' CR perception (*x*) and their loyalty towards the firm (*y*, both individual level 1) might depend on culture (level 2). Additional variables could be chosen as x or y.

(d) *Macro-to-micro-effect and interactions*, where *Z* has a direct effect on *x* and on the relation between *x* and *y*. This model type combines a cross-level effect and an interaction. For example, a researcher might want to model national culture as simultaneously determining consumers' perception of CR and the relation between CR and consumer loyalty.

Subsequently, we focus on model types (a), (c), and (d) because they are the most common and easily modifiable to generate additional model types (e.g., mediation models).

Introduction to HLM

HLMs – hierarchical linear model, random coefficient model, or variance component model – have been extensively employed (for reviews, see, e.g., Ozkaya et al. 2013). These models typically assume hierarchical data with one outcome variable that is measured at the lowest level and explanatory variables at the lowest and higher levels. One assumption of OLS regression is that the measured units are independent, i.e., that the residual e_i are uncorrelated. However, if the data are clustered, i.e., the measured units are not independent, and this clustering is not considered in a regression model (which traditional, one-level OLS regression does not do), the assumption of independence is violated. Ignoring the hierarchical structure of the data and applying traditional OLS regression tests produce underestimated standard errors. Thus, confidence intervals might be too narrow and the *p*-values too small, which produces spurious significant results. Correct standard errors will be estimated only if the variation among clusters is permitted in the analysis, which HLM allows (Kreft and Leeuw 1998, p. 2).

HLM models are often represented by a series of regression equations. As an example, we illustrate a model that allows for variation between clusters in level 1 intercepts (the function value at $x = 0$ in a linear function of x) and slopes (m in a linear function $y = mx + b$) (*random coefficient model*). Such a model is only useful for examining level 1 predictors when the data have a hierarchical structure, as is the case with a micro-level proposition in *Fig. 1*.

$$\text{Level 1:} \quad y_{ij} = \beta_{0j} + \beta_{1j} \cdot x_{ij} + r_{ij} \tag{1}$$

$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + u_{0j} \tag{2}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{3}$$

where y_{ij} is an individual i's estimate for the dependent variable y in cluster j, β_{0j} is the random intercept on level 1, β_{1j} is the random slope on level 1, x_{ij} is individual i's observation in cluster j on level 1, r_{ij} is the level 1 residual, γ_{00} is the mean intercept across all clusters, u_{0j} is the residual for the random intercept on level 2, γ_{10} is the mean slope across all clusters, and u_{1j} is the residual for the random slope on level 2. The subscript j refers to the level 2 clusters ($j = 1 \ldots J$), and the subscript i refers to the level 1 observations ($i = 1 \ldots n_j$). The residuals u_{0j} and u_{1j} are each assumed to have a mean of 0 and be independent of r_{ij}.

We can illustrate a HLM model by a firm's product managers across regions with specific levels of experience (in years) and different sales volumes of products for which they are responsible (in EUR). Considering the hierarchical structure of the data (product managers nested within regions), we may model manager i's sales (y_{ij}) on level 1 as dependent on their individual experience (x_{ij}). Thus, β_{0j} is the level 1 random intercept, i.e., the mean sales of region j; β_{1j} is the level 1 random regression slope; and r_{ij} is manager i's sales deviation from region j's mean sales. If we theoretically want to allow for variation in the intercepts and slopes across the regions, we model β_{0j} as the grand mean (i.e., average) of sales across all regions (γ_{00}) plus region j's deviation from the grand mean (u_{0j}) on level 2. β_{1j} is the average regression coefficient across all regions (γ_{10}) plus region j's deviation. The regression lines of the regression 'sales on experience' are allowed to have both a different intercept and slope in each of the regions. This model can be easily modified for the analysis of macro-to-micro effects (*means as outcomes model*) or macro-micro interactions (*intercept and slopes as outcomes model*) (see Chapter 4).

Introduction to MSEM

MSEM is a combination of HLM and SEM. Covariance-based (vs. composite-based) SEM is a well-known statistical methodology that allows one to describe the latent structure that underlies a set of manifest variables and to estimate and model relationships between latent variables (e.g., Jöreskog 1969; Kline 2011). Latent variables (factors) are constructs that cannot be directly observed but rather need to be estimated based on a number of manifest variables (Brown 2015). For example, intentional loyalty of a consumer cannot be directly measured because it resides within a consumer's mind (e.g., Oliver 2015, pp. 453–455). Typically, scholars impose the structure of a hypothesized model on the data in a sample and then test how well the observed data fit this restricted structure. Because a perfect fit between the hypothesized model and the observed data is unlikely, a differential (measurement error) naturally exists between the two (Byrne 2012). SEM was first adopted in psychology (for a review, see, e.g., Breckler 1990) but soon became increasingly popular in marketing (e.g., Bagozzi 1977) and offers measurement-design-related advantages relative to OLS regression (e.g., Vernon and Eysenck 2007):

- SEM allows a greater variety of types of indicators that can be used in the model, such as dichotomous, ordinal, categorical, and count variables.
- SEM allows for the estimation of latent constructs that consider the measurement error associated with each item instead of averaged items or factor scores of a scale (the common variance between the items is used to define the construct).
- SEM allows one to model more complicated models than OLS regression (e.g. SEM permits one to include multiple mediators rather than only one dependent variable in a model; mediation models can be simultaneously estimated without additional plugins such as the PROCESS plugin in SPSS; Hayes 2013).
- SEM offers advanced means for treating missing data, such as full-information maximum likelihood (FIML) estimation and multiple imputation.
- SEM provides exact (e.g., chi-square statistic χ^2) and approximate fit indices (e.g., comparative fit index (CFI), root mean square error of approximation (RMSEA)). Fit indices statistically indicate how well

a model fits the data compared with unconstrained/alternative models.

- SEM allows one to detect measurement invariance (MI), i.e., to determine whether or not measurement operations yield measures of the same attribute. Testing for MI is important when the data were obtained from individuals from different groups, e.g., countries or cultures (in OLS only recently proposed by Henseler et al. 2016).

Based on these advantages, the application of (single-level) SEMs with multiple indicators of individual level constructs is pervasive. Additionally, MLM studies that employ manifest variables (i.e., HLM) are common. However, progress in integrating these two dominant methodologies into a single framework has been slow. Early statistical developments laid the foundation for crucial advances, but they were difficult to implement in existing multilevel software (e.g., McDonald 1994). Recently, scholars enabled the application of MSEM, which incorporates the advantages of SEM into HLM (e.g., Marsh et al. 2009). Such a synthesis of both methods is required "when the units of observation form a hierarchy of nested clusters and some variables of interest cannot be measured directly but are measured by a set of items or fallible instruments" (Rabe-Hesketh et al. 2004, p. 168). MSEM in its current state has several advantages relative to HLM:

- HLM can only be used to model the effects of higher-level predictor variables on level 1 outcomes and the effects of level 1 predictor variables on level 1 outcomes but not the effects of lower-level variables on higher-level outcomes. In contrast, the outcome variable can be situated at any level with MSEM (Preacher et al. 2010).
- In HLM, latent variables are impossible to include or require factor loadings of 1; consequently, measurement error is problematic. In contrast, MSEM treats latent constructs (Preacher et al. 2016). Because a majority of applications naturally include latent constructs, MSEM represents a considerable advancement over HLM (Mehta and Neale 2005).
- Whereas HLM aims to indicate the explained variance by R^2 (and additional log likelihood-based model estimations that allow for the calculation of relative fit measures such as AIC and BIC), MSEM provides a variety of fit indices of the model, e.g., chi-square coefficient, CFI, and SRMR (Heck and Thomas 2015). In MSEM, level-specific model fit indices can be computed to circumvent the problem that, in the event of a poor model fit, the level at which the model fails is unclear (Ryu and West 2009).
- In HLM, within and between effects are often conflated. If steps are taken to separate these effects, then bias arises. MSEM allows for a separate examination of these effects at different levels (Lüdtke et al. 2008; Zhang et al. 2009).
- The MSEM method has been shown to outperform HLM in two-level models in terms of the bias in parameter estimates and confidence in the interval coverage. MSEM has been demonstrated to exhibit adequate efficiency, convergence rates, and power under a variety of conditions (Preacher et al. 2011).
- Finally, MSEM allows one to test for MI of items at the individual and group levels. Similar to SEM, MSEM allows one to test for these potential differences by a comparison of the strengths of the factor loadings of each construct and across clusters (e.g., Jak et al. 2013).

3. Multilevel modelling in the literature

A literature review revealed how MSEM and HLM are used, yielding insights into future efforts. *Tab. 1* summarizes 527 empirical studies published between 1/1997 and 4/2017. A focus on the highest-ranked marketing journals (IJRM, JCP, JCR, JM, JMR, JPIM, JR, JSR, JAMS, and MS), management journals (AMJ, ASQ, JE&MS, JoM, JMS, MS, OS, and SMJ), and international journals (IMR, JIBS, JIM, and JWB) assures minimum quality. The following research strategy was used to select relevant articles (e.g., Ozkaya et al. 2013). First, journals were searched using twelve keywords: hierarchical linear modelling, HLM, multilevel, multilevel modelling, MLM, multilevel structural equation modelling, MSEM, random coefficients, random intercept, random slope, Raudenbush, RCM, and random coefficients model. This procedure produced 569 articles. Second, after excluding non-MLM or conceptual/methodological studies, 527 articles were obtained. Their systematization was two-fold: characteristic research fields (general marketing, international marketing, international management, general management, and human resource management; chosen due to the relevance of MLM) and analytical method MSEM or HLM with four model types (cross-level effect, cross-level interaction, cross-level effect and interaction).

Not surprisingly, HLM is applied more often than MSEM (504 vs. 23 studies), particularly in management research (347 studies). However, HLM is only appropriate when the effects of manifest variables on a manifest level 1 outcome variable are analysed (e.g., the effect of top management team diversity on firm performance; Nielsen and Nielsen 2013), whereas management scholars also use latent variables (e.g., perceived service orientation; Aryee et al. 2016). Similarly, in marketing research. HLM studies use both manifest variables (e.g., sales volume, store size, shopping basket size, and competitive intensity; Haans and Gijsbrechts 2011) and latent constructs. Because HLM cannot treat the latter (Preacher et al. 2016), questions arise, for example, why scholars measure latent constructs but compute mean scores and how they communicate this procedure (e.g., Martin and Hill 2015, p. 409 noted that they "combine [items] into a single well-being measure" but did not state how).

	General marketing research	International marketing research	International management research	General management research	Human resource management
MSEM *Cross-level effect*	-	- *Swoboda and Hirschmann (IMR, 2017)*	- Ngobo and Fouda (JWB, 2012)	- *Hoffman et al. (AMJ, 2011)* - *Kang et al. (JoMS, 2015)* - *Kiersch & Byrne (JLOS, 2015)* - *Lorinkova and Perry (JoM, 2014)* - *Ou et al. (ASQ, 2013)*	- Den Hartog et al. (JoM, 2013) - Jensen et al. (JoM, 2013) - Johnson et al. (JoM, 2015)[5] - Kostopoulos et al. (JoM, 2013) - Shen and Benson (JoM, 2014)
Cross-level interaction	-	- *Swoboda and Hirschmann (IMR, 2017)* - *Swoboda et al. (JAMS, 2016)* - *Swoboda et al. (MIR, 2017)*	-	- *LePine et al. (AMJ, 2016)* - *Lorinkova and Perry (JoM, 2014)* - *Zhang et al. (AMJ, 2014)*	- Den Hartog et al. (JoM, 2013) - Shen and Benson (JoM, 2014)
Cross-level effect and interaction	-	- *Zhou et al. (JAMS, 2010)*	-	-	-
Further	-	-	-	- Gielnik et al. (AMJ, 2015)[1, 5] - *Oliver et al. (LQ, 2011)*[1, 5]	-
HLM *Cross-level effect*	- Ahearne et al. (JM, 2013) - *Anderson and Salisbury (JCR, 2003)* - *Auh et al. (JSR, 2011)* - Boichuk et al. (JM, 2014)[5] - Brown (MS, 1999)[3] - *Carter and Curry (JAMS, 2013)*[4] - *Chintagunta and Lee (JAMS, 2012)*[5, 7] - *Cornil et al. (JCP, 2013)* - *de Haan et al. (IJRM, 2015)*[3] - *de Jong et al. (JM, 2004)* - Dean et al. (JM, 2016)[5] - *Evanschitzky et al. (JSR, 2011)* - Haans and Gijsbrechts (JR, 2011) - *Hall et al. (JM, 2015)* - Hensen et al. (JAMS, 2016)[3] - *Homburg et al. (JM, 2009b)* - *Homburg et al. (JAMS, 2011)* - *Inman et al. (JM, 2009)* - Jensen et al. (JPIM, 2014) - *Khare and Inman (JCR, 2006)* - *Kratzer and Lettl (JCR, 2009)* - *Larivière et al. (JMR, 2016)*[3, 5] - *Lichtenstein et al. (JR, 2010)* - *Liu-Thompkins and Tam (JM, 2013)* - *Luo et al. (JMR, 2013)* - Macé (JR, 2012) - *Mallapragada et al. (JM, 2016)* - *Maxham et al. (MS, 2008)* - *Mikolon et al. (JSR, 2015)* - *Mitra and Golder (MS, 2006)*[3, 5] - *Mittal et al. (MS, 2005)*[3, 5] - *Ramanathan and McGill (JCR, 2007)* - Schepers et al. (JSR, 2011)[3] - Voss and Seiders (JR, 2003) - Wieseke et al. (JAMS, 2008) - Wieseke et al. (JSR, 2011)[3] - *Wieseke et al. (JM, 2012)* - *Wieseke et al. (JSR, 2012)* - *Wuyts et al. (JCR, 2004)* - *Ye et al. (JAMS, 2012)* - Yu et al. (JSR, 2012)	- Craig et al. (JIM, 2005) - *Laroche et al. (JAMS, 2007)* - *Martín and Cerviño (IMR, 2011)*[4] - *Martin and Hill (JSR, 2015)* - *Pauwels et al. (IJRM, 2013)*[3, 5] - Rouziès et al. (JM, 2009)	- Arregle et al. (MIR, 2006)[3] - Arregle et al. (SMJ, 2013)[3] - Autio et al. (JIBS, 2013)[5] - Bahadir et al. (JIBS, 2015) - Cantor et al. (DS, 2015) - Cullen et al. (AMJ, 2004) - Cross and Sproull (OS, 2004) - Crossland and Hambrick (SMJ, 2011)[3] - Gooderham et al. (JIBS, 2015) - Gooderham et al. (JoM, 2015)[4] - Greve et al. (JWB, 2015)[5] - Haas and Cummings (JIBS, 2015) - Handley and Gray (DS, 2015) - Hartmann and Uhlenbruck (JWB, 2015) - Hernández and Nieto (JWB, 2015)[3] - Hillman and Wan (JIBS, 2005)[3,4] - Hirst et al. (JoM, 2015) - Hitt et al. (AMJ, 2000) - Hitt et al. (OS, 2004) - Hong and Lee (JWB, 2015) - Huang and Wang (DS, 2013)[5] - Kukenberger et al. (JoM, 2015) - Kull et al. (DS, 2012) - Kwon et al. (JIBS, 2016) - Lederman (JIBS, 2010) - Luo (OS, 2007) - Luo et al. (JIBS, 2009)[3] - Martin et al. (AMJ, 2007) - Muethel et al. (JIBS, 2011) - Muethel and Bond (JIBS, 2013) - Nguyen et al. (JWB, 2013) - Parboteeah and Cullen (OS, 2003) - Parboteeah et al. (JWB, 2004) - Parboteeah et al. (JIBS, 2008) - Ping et al. (JIBS, 2004) - Sahaym and Nam (IBR, 2013) - Smale et al. (JIBS, 2015)[3] - Sun et al. (JWB, 2014) - Stephan et al. (JIBS, 2015) - Tong et al. (SMJ, 2014) - Williams and Grégoire (JIBS, 2015) - Yang et al. (JIBS, 2012)	- Ahearne et al. (SMJ, 2014) - Bamiatzi et al. (SMJ, 2016)[3] - Barsade and O'Neill (ASQ, 2014)[3, 5] - Bizzi (JoM, 2013) - Bledow et al. (AMJ, 2013) - Bloom and Milkovich (AMJ, 1998) - Cerne et al. (AMJ, 2014) - Chandrasekaran et al. (JoOM, 2012) - Choi and Shepherd (JoM, 2004, 2005) - Conlon et al. (AMJ, 2004) - Connelly et al. (SMJ, 2016) - David et al. (SMJ, 2007) - DeHoratius and Raman (MS, 2008)[4] - *Detert and Burris (AMJ, 2007)*[5] - Eggers (SMJ, 2014) - Fanelli et al. (OS, 2009) - Fedor et al. (DS, 2003) - Galaskiewicz et al. (ASQ, 2006)[7] - Gamache et al. (AMJ, 2015)[5] - Ghosh et al. (OS, 2014)[5] - Gong et al. (AMJ, 2013) - Gruys et al. (JoM, 2008)[5] - Hale Jr. et al. (AMJ, 2016)[5] - *Hill et al. (OS, 2012)*[5] - Hillman et al. (OS, 2011)[3] - Hoegl et al. (DS, 2003) - Hofmann and Stetzer (AMJ, 1998) - *Hoobler et al. (JoM, 2014)* - *Huang and Paterson (JoM, 2014)* - *Humborstad and Giessner (JoM, 2015)*[5] - *Jansen et al. (JoMS, 2016)* - Kassinis and Vafeas (AMJ, 2006) - Kim et al. (JoM, 2014)[3] - *Lam et al. (AMJ, 2015)* - *Li et al. (JoM, 2014)* - *Liu et al. (JoMS, 2012)* - *Lorinkova et al. (AMJ, 2013)*[5] - Misangyi et al. (SMJ, 2006)[3] - Mollick (SMJ, 2012)[4] - Naveh and Erez (MS, 2004)[5] - Pil and Leana (AMJ, 2009)[3] - Rust et al. (IJRM, 2016)[5]	- Aryee et al. (JoM, 2016) - Bal et al. (JoMS, 2013) - Barsade (ASQ, 2002)[3] - Beus and Whitman (JoM, 2015) - Bloom (AMJ, 1999) - Bloom and Michel (AMJ, 2002) - Bunderson (AMJ, 2003) - Cadsby et al. (AMJ, 2007) - Carlson et al. (JoM, 2011) - Chen et al. (AMJ, 2013) - Chuang et al. (JoM, 2016) - Chung et al. (AMJ, 2015) - Clark et al. (JoM, 2016) - Crocker and Eckardt (JoM, 2014) - Deadrick et al. (JoM, 1997)[5] - DiTomaso et al. (ASQ, 2007) - Dong et al. (AMJ, 2014) - Felps et al. (AMJ, 2009) - Firth et al. (AMJ, 2014)[5] - Fong et al. (SMJ, 2010) - Guillaume et al. (AMJ, 2014)[3, 5] - Griffin (JoM, 1997)[3, 5] - Harrison and Wagner (AMJ, 2016) - Hoobler et al. (AMJ, 2009) - Kammeyer-Mueller et al. (JoM, 2013) - Kauppila (JoMS, 2014)[3] - Klaas et al. (JoM, 2005) - Klein et al. (AMJ, 2004)[5] - Lau and Murnighan (AMJ, 2005) - Lee et al. (OS, 2015)[3] - Lee et al. (JoM, 2016)[5] - Levin et al. (OS, 2011) - Liao and Chuang (AMJ, 2004) - Liden et al. (JoM, 2004) - Liu et al. (AMJ, 2012a)[3] - Mannes (DS, 2009)[3] - Marrone et al. (AMJ, 2007) - Mathieu and Schulze (AMJ, 2006)[5] - Mero et al. (JoM, 2014) - Naumann and Bennett (AMJ, 2000) - Nielsen and Nielsen (SMJ, 2013)[3] - Pak and Kim (JoM, 2016) - Pearce and Xu (OS, 2012)

Tab. 1: Literature review on MLM

Tab. 1: Literature review on MLM

	General marketing research	International marketing research	International management research	General management research	Human resource management
			- Zhang et al. (JWB, 2016)[4] - Zheng et al. (JIBS, 2013)	- Sauerwald et al. (JoM, 2016)[3] - *Schaubroeck et al. (AMJ, 2012)* - Shepherd et al. (AMJ, 2013) - Short et al. (SMJ, 2007)[3] - *Song et al. (JoM, 2009)* - Sutcliffe and McNamara (OS, 2001) - *Sparrowe et al. (AMJ, 2006)* - Spencer and Gomez (SMJ, 2011)[4] - Tatarynowicz et al. (ASQ, 2015)[3, 5] - *Waldman et al. (JoM, 2015)* - Wallace et al. (JoM, 2013) - Yang and Schwarz (OS, 2016)[5] - *Zhang et al. (AMJ, 2015)*[3]	- Ployhart et al. (AMJ, 2006)[3] - Ployhart et al. (AMJ, 2009)[5] - Quinn and Bunderson (JoM, 2016) - Scott et al. (AMJ, 2012)[3, 5] - Seibert et al. (AMJ, 2004) - Trougakos et al. (AMJ, 2014) - Wang et al. (OS, 2016) - Wiltermuth and Flynn (AMJ, 2013) - Vancouver (JoM, 1997) - van der Vegt et al. (AMJ, 2006)
Cross-level interaction	- Ahearne et al. (JM, 2013) - Ahearne et al. (JAMS, 2013) - *Aksoy et al. (JSR, 2011)*[5] - *Auh et al. (JSR, 2011)* - Behrens et al. (JPIM, 2014) - Boichuk et al. (JR, 2013) - Boichuk et al. (JM, 2014)[5] - *de Jong et al. (JM, 2004)* - *Groening et al. (JMR, 2016)* - Hohenberg and Homburg (JM, 2016) - *Homburg et al. (JM, 2009a[5]; 2009b)* - *Homburg et al. (JAMS, 2010a, b, 2011)* - Hulland et al. (JAMS, 2012)[3] - Karniouchina et al. (JM, 2011)[4, 5] - Klaukien et al. (JPIM, 2013) - Krasnikov et al. (JM, 2009)[5] - Kraus et al. (JR, 2015) - *Kwon and Nayakankuppam (JCR, 2015)* - *Larivière et al. (JMR, 2016)*[3;5] - *Lembregts and Pandelaere (JCR, 2013)* - *Ma and Dubé (JM, 2011)* - Menguc et al. (JM, 2016)[5] - *Mikolon et al. (JSR, 2015)* - *Mourali et al. (JCR, 2007)* - *Netemeyer et al. (JR, 2012)* - Palmatier et al. (MS, 2006) - Patzelt et al. (JPIM, 2011) - *Petersen et al. (JM, 2015)* - *Pham et al. (IJRM, 2013)*[3] - *Rapp et al. (JAMS, 2013*[3]; 2015) - Rapp et al. (JR, 2015) - *Richins (JCR, 2013)*[5] - Schmitz (JAMS, 2013) - Schmitz et al. (JM, 2014) - *Sridhar and Srinivasan (JM, 2012)* - Thomaz and Swaminathan (JM, 2015)[3] - *van Birgelen et al. (IJRM, 2002)* - *van Dolen et al. (JR, 2002)* - *van Vaerenbergh et al. (JSR, 2012)*[5] - Venkatesan et al. (JR, 2007) - Vomberg et al. (SMJ, 2015) - Wieseke et al. (JAMS, 2008)	- Akdeniz and Talay (JAMS, 2013) - *Deephouse et al. (JWB, 2016)* - Du and Choi (JIBS, 2010) - *Echambadi et al. (JR, 2013)*[3, 4] - *Farley et al. (IJRM, 2004)* - *Frank et al. (JAMS, 2014)*[4] - Griffith et al. (JIM, 2014) - Griffith and Rubera (JIM, 2014) - Hewett and Krasnikov (JIM, 2016)[3] - *Hsieh et al. (JAMS, 2004)*[3] - Lund et al. (JIM, 2013) - Magnusson et al. (JIM, 2009)[4] - *Martin and Hill (JCR, 2012)* - *Martin and Hill (JSR, 2015)* - *Möller and Eisend (JIM, 2010)* - Qiu (JIM, 2014) - *Schumann et al. (JIM, 2010)* - *Steenkamp and Geyskens (JM, 2006)* - *Strizhakova and Coulter (JIM, 2015)* - *Walsh et al. (JIM, 2014)*	- Alexander (MIR, 2012) - Auh et al. (JAMS, 2015) - Aulakh et al. (JIBS, 2013)[4] - Ault (JIBS, 2016)[5] - Bahadir et al. (JIBS, 2015) - Fragale et al. (ASQ, 2012) - Goerzen et al. (JIBS, 2013) - Goldszmidt et al. (JBR, 2011)[4, 5] - Haas and Cummings (JIBS, 2015) - Hillman and Wan (JIBS, 2005)[3,4] - Hirst et al. (JoM, 2015) - Hon und Lu (JWB, 2015) - Huang and Wang (DS, 2013)[5] - Jiang et al. (JIBS, 2011) - Ju et al. (JIM, 2013)[3] - Ketelhöhn and Quintanilla (JBR, 2011) - Kirkman et al. (AMJ, 2009) - Kwon et al. (JIBS, 2016) - Lam et al. (JIBS, 2012) - Levin and Barnard (JIBS, 2013) - Li et al. (JIBS, 2011) - Luo and Bu (JWB, 2016) - Mani et al. (SMJ, 2007)[4] - Nam et al. (JoIM, 2014) - Nielsen and Nielsen (JWB, 2011) - Nguyen et al. (JWB, 2013) - Paris et al. (JIBS, 2009) - Power et al. (DS, 2015)[4] - Raab et al. (JWB, 2014) - Rubera et al. (JPIM, 2012) - Setia and Speier-Pero (DS, 2015) - Tian and Slocum (JWB, 2015) - Wu (IJRM, 2013) - Yang et al. (JIBS, 2012) - Young and Makhija (JIBS, 2014) - Zacharakis et al. (JIBS, 2007)	- Ahearne et al. (SMJ, 2014) - Bacharach and Bamberger (AMJ, 2007) - Bamiatzi et al. (SMJ, 2016)[3] - Banalieva et al. (SMJ, 2015)[5] - Bhave and Glomb (JoM, 2016) - Bloom and Milkovich (AMJ, 1998) - Butts et al. (AMJ, 2015) - Chowdhury and Endres (AMJ, 2010) - Cuervo-Cazurra and Dau (AMJ, 2009)[3] - Garrett et al. (JPIM, 2013) - Ghosh et al. (OS, 2014)[5] - Hough and White (SMJ, 2003)[4] - Jansen et al. (SMJ, 2012) - *Jansen et al. (JoMS, 2016)* - *Lam et al. (JoM, 2015)* - *Liu et al. (AMJ, 2012b)* - *McClean et al. (AMJ, 2013)*[5] - Morrison and Vancouver (JoM, 2000) - Murnieks et al. (JoMS, 2011) - Pil and Leana (AMJ, 2009) - Sauerwald et al. (JoM, 2016)[3] - *Schaubroeck et al. (AMJ, 2013)*[3] - Shepherd and Patzelt (JoMS, 2015) - Spencer and Gomez (SMJ, 2011)[4] - Steven et al. (JoOM, 2014) - van der Vegt et al. (AMJ, 2005) - *Waldman et al. (JoM, 2015)* - Wallace et al. (JoM, 2013) - Wang et al. (AMJ, 2011)[3] - Wood and Williams (JoMS, 2014) - Yang and Schwarz (OS, 2016)[5] - Zhu et al. (AMJ, 2016)	- Aryee et al. (JoM, 2016) - Bal et al. (JoMS, 2012) - Bommer et al. (AMJ, 2007) - Courtright et al. (AMJ, 2016) - Daniels et al. (JoM, 2013)[3] - Diestel et al. (AMJ, 2014) - Duffy et al. (AMJ, 2012) - Dwertmann and Boehm (AMJ, 2016) - Ferris et al. (JoM, 2012)[5] - Georgakakis and Ruigrok (JoMS, 2017)[3] - Glaser et al. (AMJ, 2016) - Gruys et al. (JoM, 2008)[5] - Hirst et al. (AMJ, 2009; 2011) - Hon et al. (JoM, 2014) - Ilies et al. (AMJ, 2006, 2009[5]) - Judge et al. (JoAP, 2006) - Joshi et al. (AMJ, 2006)[3] - Joshi and Knight (AMJ, 2015)[5] - Kammeyer-Mueller et al. (JoM, 2013) - Kauppila (JoMS, 2014)[3] - Klaas et al. (JoM, 2005) - Kidwell et al. (JoM, 1997) - Liao and Chuang (AMJ, 2004) - Liu et al. (AMJ, 2012a)[5] - Marrone et al. (AMJ, 2007) - Mathieu and Schulze (AMJ, 2006)[5] - Messersmith et al. (OS, 2014)[5] - Nielsen and Nielsen (SMJ, 2013)[3] - Park et al. (JoM, 2015) - Parker et al. (AMJ, 2013) - Peng et al. (JoMS, 2015) - Ployhart et al. (AMJ, 2006)[3] - Scott and Barnes (AMJ, 2011) - Scott et al. (AMJ, 2012)[3;5] - Scott and Judge (JoM, 2006) - Shin et al. (AMJ, 2012)[1;3, 5] - Spell and Arnold (JoM, 2007) - Sun et al. (AMJ, 2007) - van der Vegt et al. (JoM, 2000) - Wanberg et al. (AMJ, 2012) - Whitener (JoM, 2001) - Wilk and Makarius (OS, 2015)

Tab. 1: Literature review on MLM

	General marketing research	International marketing research	International management research	General management research	Human resource management
	- Wieseke et al. (JM, 2009)[1;3] - *Wieseke et al. (JSR, 2012)*				- Yeo and Neal (JoM, 2013)
Cross-level effect and interaction	- Auh et al. (JAMS, 2014) - *van Dolen et al. (JR, 2007)*	- Magnusson et al. (IMR, 2014) - *Steenkamp et al. (JM, 1999)*	- Cerne et al. (JoIM, 2013) - Hui et al. (JIBS, 2004) - Joshi et al. (OS, 2009) - Nguyen et al. (JWB, 2013) - Ralston et al. (JIBS, 2009)	- *DeCelles et al. (OS, 2013)* - *Leroy et al. (JoM, 2012)* - Magoshi and Chang (JWB, 2009)	- Chang (JWB, 2006) - Wang et al. (OS, 2013)
Further	- Balka et al. (JPIM, 2014)[1] - *Berger and Schwartz (JMR, 2011)*[1] - *Blut et al. (JR, 2014)*[6] - *Calli et al. (IJRM, 2012)*[1, 5] - *Chatterjee et al. (MS, 2003)*[1] - Edeling and Fischer (JMR, 2016)[1, 6] - *Eisend (JR, 2014)*[1, 3, 6] - *Elsen et al. (JMR, 2016)*[1] - *Fernbach et al. (JCR, 2015)*[1, 5] - *Homburg et al. (JAMS, 2010c)*[1] - *Janakiraman et al. (JR, 2016)*[6] - *Keiningham et al. (JSR, 2015)*[1, 4] - Köhler et al. (JMR, 2016)[6] - Krasnikov and Jayachandran (JM, 2008)[6] - *Liu (JM, 2007)*[1, 3, 5] - *Lovett et al. (JMR, 2013)*[1] - McFarland et al. (JAMS 2016)[1] - *Meißner et al. (JMR, 2016)*[1] - *Nagengast et al. (JR, 2014)*[1] - *Neumann and Böckenholt (JR, 2014)*[6] - *Ou et al. (JSR, 2013)*[6] - Palmatier (JM, 2008)[1] - Rouziès et al. (JAMS, 2014)[1] - Rubera and Kirca (JM, 2012)[6] - Sarin et al. (JMR, 2012)[1] - Troy et al. (JM, 2008)[6] - *Watson et al. (JAMS, 2015)*[6] - Wei et al. (JPIM, 2013)[1] - You et al. (JM, 2015)[6] - *Zlatevska et al. (JM, 2014)*[6]	- *Boer and Fischer (PB, 2013)*[6] - Eisend et al. (JIM, 2016)[6] - *Knoll and Matthes (JAMS, 2017)*[3, 6] - Leenders and Eliashberg (IJRM, 2011)[1] - *Magnusson et al. (IMR, 2011)*[1, 4] - *Pick and Eisend (JIM, 2016)*[6] - *Schlager and Maas (JIM, 2013)*[1] - Wieseke et al. (JM, 2009)[1, 3]	- Arregle et al. (JIBS, 2016)[1, 3] - Broderick et al. (JIBS, 2007)[1] - Castellaneta and Gottschalg (SMJ, 2014)[1, 4] - Fischer and Mansell (JIBS, 2009)[3, 6] - Judge et al. (SMJ, 2014)[1] - Levy et al. (JIBS, 2014)[1] - Liu et al. (ASQ, 2012)[1] - Mäkelä et al. (JIBS, 2013)[1] - Marano et al. (JoM, 2016)[5, 6] - Meyer and Su (JWB, 2015)[1] - Queenan et al. (DS, 2016)[1] - Shao et al. (JIBS, 2013)[1] - Steel and Taras (JoIM, 2010)[6] - Tröster and van Knippenberg (JIBS, 2012)[1] - van Essen et al. (JIBS, 2012)[1, 6]	- Arrfelt et al. (AMJ, 2013)[1] - Bridwell-Mitchell and Lant (OS, 2014)[1, 3, 4] - *Carton and Rosette (AMJ, 2011)*[1] - Cobb et al. (AMJ, 2016)[1] - Connelly et al. (SMJ, 2012)[1] - *deRue et al. (OS, 2015)*[1] - Gabriel and Diefendorff (AMJ, 2015)[1, 7] - Hough (SMJ, 2006)[2, 3, 4] - Hüffmeier et al. (JoM, 2014)[6] - Jayachandran et al. (SMJ, 2013)[1] - Karniouchina et al. (SMJ, 2013)[2, 4, 5] - Leana et al. (AMJ, 2009)[1] - Lo et al. (JoOM, 2013)[1] - Majumdar and Bhattacharjee (OS, 2014)[1] - *Marinova et al. (OS, 2013)*[1] - Marquis et al. (OS, 2016)[1] - Marr and Cable (AMJ, 2014)[1] - McNamara et al. (SMJ, 2003)[1] - Nair et al. (JoOM, 2013)[1, 3] - Naveh and Marcus (JoOM, 2005)[1, 5] - *Quigley and Hambrick (SMJ, 2015)*[2, 3] - *Quigley and Graffin (SMJ, 2016)*[2] - Walter et al. (OS, 2015)[1] - Wu et al. (JoM, 2011)[1, 4]	- Abbas et al. (JoM, 2012)[1] - Barnes et al. (JoM, 2011)[1] - Beck and Schmidt (JoM, 2015)[1] - Bono et al. (AMJ, 2013)[1, 5] - Boxall et al. (JoMS, 2011)[1] - Callahan et al. (JoM, 2003)[6] - Chen et al. (AMJ, 2011)[1, 5] - Collins and Mossholder (JoM, 2014)[1] - Dalal et al. (AMJ, 2009)[1] - de Stobbeleir et al. (AMJ, 2011)[1] - Galunic et al. (AMJ, 2012)[1] - Gong et al. (AMJ, 2009)[1] - Gong et al. (JoM, 2012)[1, 3] - Jehn et al. (AMJ, 2010)[1] - Kilduff et al. (OS, 2016)[1, 5] - Lount Jr. and Wilk (MS, 2014)[1] - Menon and Phillips (OS, 2011)[1] - Owens et al. (OS, 2013)[1] - Peng et al. (AMJ, 2014)[1] - Rothbard and Wilk (AMJ, 2011)[1, 3] - Scott et al. (AMJ, 2015)[1, 5] - Shin et al. (AMJ, 2012)[1, 3, 5] - Sluss et al. (AMJ, 2012)[1] - Strauss and Parker (JoM, 2015)[1, 5] - Sturman et al. (JoM, 2003[6]; 2008)[1, 4, 5] - Thatcher and Greer (JoM, 2008)[1] - Uy et al. (AMJ, 2015)[1] - Vardaman et al. (OS, 2015)[1] - Vinhas et al. (MS, 2012)[1] - Walker et al. (AMJ, 2013)[1, 5] - Wallace et al. (JoM, 2016)[6]

Notes: The studies are not included in the reference list. Studies in italics: marketing = consumer-research-based; general management = leadership; [1]micro-level propositions (i.e., controlling for hierarchical data structure); [2]variance partitioning only; [3]three or more levels; [4]cross-classified data structure; [5]longitudinal analysis; [6]meta-analysis; [7]hierarchical growth model.

Most studies analysed *cross-level effects* (212 studies) and *cross-level interactions* (189 studies); only ten studies and eight studies, respectively, applied MSEM. For example, Magnusson et al. (2014) analysed the effect of masculinity and individualism on sales collaboration. Sridhar and Srinivasan (2012) employed cross-level interactions to investigate how other consumers' online product ratings moderate the effect between product failure and a reviewer's online product rating. *Cross-level effects and cross-level interactions* were used in 15 studies (ten studies simultaneously and five studies successively); only one study applied MSEM but with a dichotomous level 2 variable (Zhou et al. 2010). *Additional model types* were applied – in combination – in 111 studies (of which two were MSEM): *micro-level propositions* (85 studies), *longitudinal data* (70 studies), *cross-classified models* (27 studies), *variance partitioning* (four studies), or *hierarchical growth models* (three studies). Twenty-four studies were performed using *meta-analyses*. Finally, 456 studies included two levels, and 71 studies included three or more levels (all HLM), e.g., customers nested in firms that were nested in industries (Larivière et al. 2016).

In the marketing research, 71 studies used *cross-level interactions* (one MSEM), 48 studies used *cross-level effects* (three MSEM), five used *cross-level effects and interactions* (one MSEM), and 38 studies used *additional model types* (e.g., Calli et al. 2012 conducted a longitudinal analysis to analyse whether advertising effects vary with the hour of the week). Ninety-seven of the 162 studies addressed consumers as the level 1 unit of analysis, e.g., explored the effects of employee and customer empathy on customer loyalty (Wieseke et al. 2012) or how the length of the customer relationship moderates the effect of customer-company identification on customers' willingness to pay (Homburg et al. 2009). The remaining 65 studies addressed various topics, e.g., eleven sales management, nine B2B marketing, and seven product development studies.

Concerning MSEM, only 23 studies were found (18 studies in management research and five studies in marketing research), with the following model types:

- *Cross-level effects* (twelve studies): For example, Ngobo and Fouda (2012) investigated country-level good governance as drivers of firm-level performance. Performance was measured using return on sales, whereas country governance was measured as a construct of three dimensions (authority selection and replacement, government capacity, and respect for institutions). Kiersch and Byrne (2015) analysed the effect of the group-level construct authentic leadership on the employee-level outcomes stress, turnover intention, and organizational commitment and measured latent constructs. Jensen et al. (2013) investigated how high-performance work systems in departments influence employees' perceptions of such systems.
- *Cross-level interactions* (eight studies): For example, Swoboda et al. (2017) tested institutional country distances and firm-specific resources as moderators of the individual-level CR-loyalty relationship. LePine et al. (2016) analysed how unit-level charismatic leadership traits moderate the employee-level effects between challenge stressors, challenge appraisals, and task performance. Finally, Shen and Benson (2016)

Source	**Description**
Textbooks on/including MSEM	
Byrne (2012)	SEM with a chapter on MSEM.
Finch and Bolin (2017)	MLM with a chapter on MSEM.
Heck and Thomas (2015)	MLM with a chapter on MSEM.
Hox (2010)	MLM with a chapter on MSEM.
Kelloway (2015)	SEM with a chapter on MSEM.
Papers/chapters (general)	
Cheung and Au (2005)	MSEM in cross-cultural research demonstrating Muthén's (1994) MSEM procedure.
du Toit and Du Toit (2008)	Explaining a general two-level SEM method.
Hox (2013)	Book chapter on MSEM that gives an easy to read introduction to the topic in general.
Mehta and Neale (2005)	Using concepts that are well understood in the context of SEM.
Muthén (1994)	Technique for multilevel covariance structure modelling with latent variables.
Rabe-Hesketh et al. (2007)	Book chapter on a technique for handling latent variables at different levels.
Rabe-Hesketh et al. (2012)	Book chapter that gives a methodological overview of MSEM.
Selig et al. (2008)	Book chapter that compares latent variable SEM between multigroup and multilevel approaches.
Papers (specific issues)	
Geldhof et al. (2014)	Methods to compute multilevel reliability.
Jak et al. (2013)	Methods to compute multilevel measurement invariance.
Newsom (2002)	Technique for handling dyadic data in MSEM.
Preacher et al. (2010)	MSEM framework for assessing multilevel mediation.
Preacher et al. (2011)	Empirical evidence of MSEM's advantages over HLM for multilevel mediation.
Preacher et al. (2016)	Technique for handling moderation within and across levels in MSEM.

Tab. 2: Methodological MSEM literature

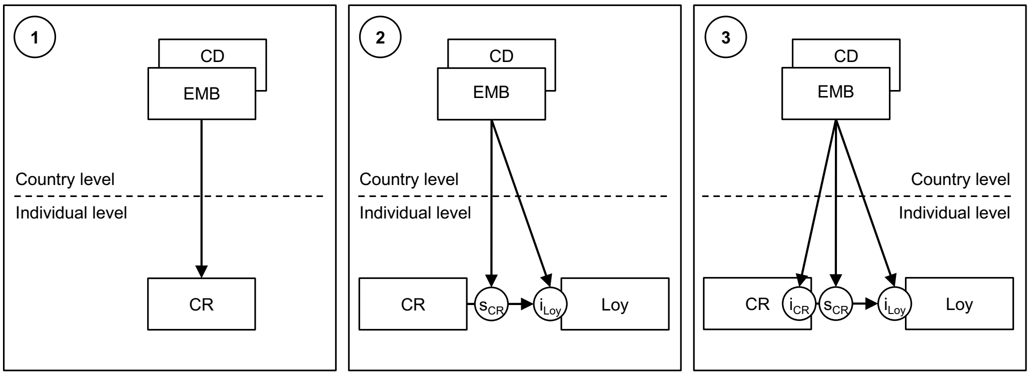

Fig. 2: Conceptual framework of example models

Note: CR = Corporate reputation; Loy = Loyalty; EMB = Embeddedness; CD = Country development; s_{CR} = Slope of CR on loyalty; i_{Loy} = Intercept of loyalty; i_{CR} = Intercept of CR; controls on individual level: age, gender; control on country level: group size.

analysed how organizational cooperative norms moderate the relationship between an individual's organizational identification and their helping behaviour using variables on the organizational/employee levels.

- *Cross-level effects and cross-level interactions* were used only by Zhou et al. (2010), who observed both effects simultaneously when analysing how foreign vs. local brand origin directly affects perceived brand value and moderates the interaction between perceived brand foreignness, confidence in brand origin identification, and brand value.
- *Further model types* (two studies): C*ontrolled for a hierarchical data structure* (e.g., resulting from measurement at multiple time points) but measured all constructs at the individual level. Gielnik et al. (2015) investigated whether and how entrepreneurial effort predicts changes in entrepreneurial passion. Oliver et al. (2011) analysed the relationships among positive family functioning, adolescent self-concept, and transformational leadership.

In summary, the relatively few MSEM studies were published from 2010–2017, which is consistent with the previously mentioned statistical developments that enabled the computation of MSEM. Therefore, *Tab. 2* lists methodological literature about MSEM and provides readers with an easy reference: textbooks that address MSEM (Byrne 2012 provides a good introduction; the book by Heck and Thomas 2015 includes more advanced topics and many examples), papers about MSEM (Hox 2013 can easily be understood by beginners; Rabe-Hesketh et al. 2007 develop a Generalized Linear Latent and Mixed Modeling technique that targets advanced readers), papers about MSEM from a statistical perspective (e.g., Muthén 1994), and papers about specific issues (e.g., Geldhof et al. 2014 on reliability; Jak et al. 2013 on measurement invariance).

4. Systematic evaluation of the requirements, options, and challenges of MSEM

Throughout the illustration of a step-by-step procedure for conducting MSEM, we first describe the discussion status in the literature, i.e., the requirements, options, and challenges associated with each characteristic step of a study (sample, measurement, and method). Second, we provide insights from our experience and use a study as an example to obtain novel results for three MSEM models: cross-level effects, cross-level interactions, and cross-level effects and interactions. The management of multinational corporations' (MNCs) CR across nations provides the context. We use surveys conducted by a MNC every year in up to 40 countries to coordinate subsidiaries (e.g., communication budget). For the MNC, links to 140 additional countries of presence are possible based on national factors known as diminishing or reinforcing CR perceptions and effects in extant surveys. Two important national institutions were selected (embeddedness (EMB) and country development (CD); for further details see, e.g., Berry et al. 2010); the research questions are as follows:

- Do EMB and CD explain consumers' CR perceptions across nations, and if so, how?
- Do EMB and CD explain CR effects on consumer loyalty across nations, and if so, how?
- Do EMB and CD simultaneously explain both CR perceptions and effects on consumer loyalty across nations, and if so, how?

The relationships are theoretically likely. We know, for example, that a strong EMB of societies positively affects consumers' use of important signals in decision situations, such as firms' reputation, and that EMB particularly affects CR effects across nations (e.g., Swoboda et al. 2016). The research questions imply a hierarchy with two levels (see *Fig. 2*). At the higher level (level 2), EMB and CD represent variables known from institutional theories, whereas at level 1, consumer CR perceptions and loyalty are present. The variables at level 1 are nested and affected by the level 2 variables. The models differ considerably: model 1 examines culturally bounded consumer perception differences across nations, whereas model 2 examines culture as a moderator of CR effect differences, and model 3 simultaneously examines culture as both an antecedent and moderator.

4.1. Sample

Sample size and data quality

Sample size and data quality are critical for MLM. HLM is less demanding in terms of sample size, whereas MSEM requires a considerably larger number of level 2 clusters. Although theory should primarily guide the number of clusters chosen, scholars have not yet identified the statistically required sample size in MSEM. Two requirements are discussed.

- The relationship of sample size with level 2 and level 1, e.g., the total number of clusters and the sample size on the individual level (in two-level models). This relationship is essential as MSEM models are confirmatory factor analysis (CFA) models that are restricted to fit to cluster-level covariances instead of individual-level covariances (Mehta and Neale 2005). For HLM, simulations indicate that more than 30 groups and fewer than 30 observations per group are needed to capture the effects of level 2 variables (Maas and Hox 2005). For MSEM, similar guidelines are pending.
- Small sample sizes on level 1 are not problematic (e.g., Atkins 2005). As long as other groups are larger, even group sizes of one are considered to be technically possible (Snijders and Bosker 2012, p. 56; Newsom 2002 proposes an approach to MSEM with dyadic data), with limitations (e.g., models with a random intercept and one random slope are just identified; Hox 2013, p. 290). However, the level 2 number of clusters has not been discussed. Two issues are critical: whether the computations run and converge at all and whether the results are unbiased. Unfortunately, scholars provide sweeping recommendations, particularly for level 2 groups. For example, for HLM, Hox (2010) recommended 100 groups for large models and 50 groups for smaller models. Our experience indicates that the number of level 2 clusters needs to be larger than the number of parameters to be estimated. A simulation study by Maas and Hox (2004) reported standard errors for fixed effects to be slightly biased with less than 50 groups. However, MSEM exists with fewer groups, e.g., cross-level interaction models with 37/40 countries and stable results in revised models over different years of analysis (e.g., Swoboda and Hirschmann 2017; Swoboda et al. 2016 with five/three items for independent/dependent variables on level 1, several controls, and up to seven moderators) or cross-level effect models with 37 teams (e.g., Kiersch and Byrne 2015).

Options to address this evolving issue are two-fold. First, in our initial studies and after contact with Muthén and Muthén (1998–2017), we pragmatically attempted to determine how large the number of clusters needs to be for computations to run to test a theoretically deduced level 1 model using a given sample. Swoboda et al. (2017) included 43 countries from surveys in two years to perform MSEM. Second, power analyses should be employed in future studies to determine the required number of clusters a priori. Monte Carlo simulation studies can be employed (e.g., Mathieu et al. 2012; Muthén 2002; Muthén and Muthén 1998–2015, p. 407). Spybrook et al. (2011) proposed a specific software in MLM applications for this purpose.

A challenge concerning sample size in MSEM is that high-level sample sizes are typically smaller than low-level sample sizes and are too small. In cross-cultural research, for example, the number of important countries is limited and might force scholars to switch to HLM, as indicated in the literature review. Furthermore, scholars chose a balanced sample number on level 1. Swoboda et al. (2016) excluded countries with small samples and used an average of 355 respondents in the five smallest countries to randomly reduce the number of respondents across 40 countries, which ranged between 280 and 1,023. They also excluded countries with smaller samples ($n < 154$). Although the authors compared the initial sample and the reduced sample, this procedure is not necessary in MSEM because MSEM considers varying cluster sizes.

In our example, we use a study performed in cooperation with a German MNC operating in the pharmaceutical/chemical industry, where CR is particularly important. The MNC centrally controls CR perceptions in important countries each year by surveying a maximum of 1,000 consumers per country via existing panels (based on reasonable screening criteria).[2] We merge data from 51 countries from 2012 to 2013 to obtain a sufficiently large sample (see *Tab. 3*; n = 145 to 976).

Data set-up and initial analyses

The procedures for data set-up and the initial analyses in MSEM are rather traditional. Although methods for detecting multilevel outliers have been statistically discussed (e.g., Ieva and Paganoni 2015), they have not been employed in studies. Therefore, outlier diagnostics can be based on the Mahalanobis distance (Kline 2011, p. 54; reducing the number of respondents from 28,881 to 26,897 in our example). Testing representativeness can be based on comparisons with official demographics data (our quota sampling is not representative; see *Tab. 3*). Traditional tests for univariate and multivariate normality follow (using Mardia's coefficient; all values indicated that the data were normally distributed). In the case of non-normality, MSEM may use the robust maximum likelihood (MLR) estimation (e.g., Heck and Thomas 2015, p. 46).

Testing for hierarchical data structure

The test for a hierarchical data structure is specific to MLM as it determines whether MLM is required. Computations of intraclass correlations (ICCs) are recommended (Barile 2016). The ICC value quantifies the amount of variance at the individual and group levels and is defined as

$$ICC = \frac{\tau^2}{\tau^2 + \sigma^2} \tag{4}$$

		Gender (%)		Age groups (years, %)				
	N	Male	Female	18 to 25	26 to 35	36 to 45	46 to 55	56 to 65
Argentina	886	48.1	51.9	20.0	31.3	26.7	22.0	0.0
Australia	815	50.2	49.8	9.4	20.4	23.4	22.2	24.5
Austria	894	51.1	48.9	13.2	22.0	24.7	22.1	17.9
Belgium	896	51.5	48.5	14.8	18.6	24.1	23.3	19.1
Brazil	821	50.8	49.2	33.1	29.6	23.1	14.1	0.0
Canada	817	49.0	51.0	19.1	20.7	22.5	17.1	20.6
Chile	425	48.5	51.5	20.2	27.3	26.1	26.4	0.0
China	663	50.2	49.8	20.2	35.3	28.1	16.4	0.0
Colombia	976	49.8	50.2	23.5	30.0	26.9	19.6	0.0
Costa Rica	445	48.8	51.2	22.9	30.3	29.0	17.8	0.0
Czech Republic	552	55.8	44.2	9.2	21.7	26.3	20.1	22.6
Denmark	750	55.2	44.8	12.0	19.1	20.8	22.5	25.6
Estonia	284	50.4	49.6	14.8	23.6	24.3	21.5	15.8
Finland	824	49.5	50.5	14.8	17.6	21.1	21.8	24.6
France	690	51.4	48.6	14.1	19.6	22.6	14.9	28.8
Germany	911	50.3	49.7	16.0	17.0	25.1	24.3	17.6
Ghana	145	57.9	42.1	30.3	33.8	22.8	13.1	0.0
Guatemala	462	55.0	45.0	13.2	40.9	30.5	15.4	0.0
India	720	51.8	48.2	22.2	26.1	23.2	28.5	0.0
Indonesia	927	49.9	50.1	25.9	26.8	21.8	25.6	0.0
Italy	893	47.6	52.4	13.2	21.6	23.3	21.7	20.2
Japan	683	49.2	50.8	9.4	16.7	23.1	23.1	27.7
Kenya	213	61.0	39.0	24.4	31.9	24.4	19.2	0.0
Latvia	315	46.0	54.0	10.2	27.0	26.7	26.3	9.8
Lithuania	359	49.9	50.1	11.1	20.1	26.5	26.5	15.9
Malaysia	757	51.8	48.2	22.3	21.8	24.2	31.7	0.0
Mexico	910	47.1	52.9	32.1	28.8	22.6	16.5	0.0
Morocco	300	50.0	50.0	26.7	32.0	23.0	18.3	0.0
Netherlands	816	50.1	49.9	11.5	17.2	26.3	23.8	21.2
New Zealand	623	51.4	48.6	10.1	16.7	25.0	26.0	22.2
Nigeria	167	49.7	50.3	29.3	34.7	19.2	16.8	0.0
Norway	557	49.2	50.8	9.7	16.2	21.7	25.9	26.6
Peru	488	49.2	50.8	25.4	31.1	24.8	18.6	0.0
Philippines	905	48.6	51.4	29.2	27.4	19.6	23.9	0.0
Poland	927	51.5	48.5	19.0	23.7	18.7	21.4	17.3
Portugal	956	49.5	50.5	15.1	27.2	31.4	26.4	0.0
Russia	677	55.8	44.2	23.2	25.3	25.4	26.1	0.0
Singapore	788	51.1	48.9	10.8	20.9	28.4	39.8	0.0
Slovakia	742	53.1	46.9	12.4	21.7	22.9	20.6	22.4
South Africa	755	49.1	50.9	25.4	25.6	19.6	29.4	0.0
South Korea	645	52.6	47.4	10.4	22.2	29.6	37.8	0.0
Spain	913	49.4	50.6	14.0	23.3	24.9	20.6	17.2
Sweden	485	55.5	44.5	13.4	13.6	19.4	22.9	30.7
Switzerland	411	48.9	51.1	13.4	19.7	25.1	23.6	18.2
Thailand	710	53.0	47.0	17.3	25.4	28.7	28.6	0.0
Turkey	893	49.2	50.8	16.2	36.8	30.1	16.8	0.0
United Kingdom	587	51.6	48.4	13.3	18.6	24.4	21.5	22.3
Ukraine	604	48.2	51.8	12.1	30.3	30.8	26.8	0.0
USA	890	50.3	49.7	16.7	19.1	24.0	23.5	16.6
Venezuela	927	50.7	49.3	25.4	31.3	25.0	18.3	0.0
Vietnam	630	51.7	48.3	22.5	29.8	26.5	21.1	0.0
Total	34.429	50.6	49.4	17.8	24.4	24.7	22.8	10.2

Tab. 3: Sample characteristics

where τ^2 is the group-level variance, and σ^2 is the individual-level variance. The ICC value ranges from 0 to 1 and represents the fraction of the variance that occurs at the group level. A large ICC value indicates a large clustering effect with little individual variability, whereas a small ICC value indicates a weak clustering effect with extensive individual variability within the groups. A specific cut-off value for ICC that requires MLM does not exist. Hox (2013, p. 282) described an ICC of 36 % as "relatively large". However, small ICCs (1 % to 5 %) can lead to bias in parameter estimates and significance tests (e.g., Hox 2010, p. 243). Julian (2001) reported simulation results in which ICCs of 5–15 % resulted in biased estimates without MLM. To the best of our knowledge, specific challenges for MSEM regarding the computation of ICCs do not exist. However, some scholars do not test for ICC (e.g., Hohenberg and Homburg 2016), whereas other scholars use MSEM and report maximum ICCs of 27 % (e.g., Zhou et al. 2010).

In our example, we address results for the cross-level interaction models (not the other two models) due to space restrictions. To test the breakdown of variance in the criterion variables, we estimated a null model that contained no predictor variables. In the cross-level interaction model, 23.3 % [.153/(.153+.505)] of the variation in CR was attributable to country differences. MLM is appropriate.

4.2. Measurements

Some requirements regarding measurements are specific to MSEM, e.g., testing for reliability/validity or MI. We address them using our example.

Measures

On level 1 (the individual level), we considered general aspects – the hierarchy of effects in the panels – and relied on established scales from previous studies (using five-point Likert-type scales; see *Tab. 4*). CR was measured by 15 items that captured its five dimensions (customer orientation, good employer, product/service quality, social/environmental responsibility, and reliability/financial strength), whereas loyalty was measured with three items (Oliver 1999; 2015, pp. 453–455; Walsh and Beatty 2007). The scales were pre-tested (e.g., in eight countries with satisfactory values for reliability/validity),

Construct		Item	MV/Std	λ
Corporate reputation (Walsh and Beatty 2007; Walsh et al. 2009)	Customer orientation (CO)	[MNC] has employees who are concerned about customer needs.	3.48/.855	.915
		[MNC] has employees who are polite to their customers.	3.49/.841	.897
		[MNC] is concerned about its customers.	3.59/.900	.886
	Good employer (GE)	[MNC] appears to be a good employer.	3.61/.865	.909
		[MNC] seems to have an excellent leadership style.	3.59/.875	.896
		[MNC] seems to treat its employees well.	3.53/.838	.909
	Product range quality (PRQ)	[MNC] is a strong, reliable company.	3.84/.901	.904
		[MNC] offers high-quality products.	3.82/.895	.914
		[MNC] develops innovative products.	3.71/.885	.874
	Social/environmental responsibility (SER)	[MNC] would reduce its profits to ensure a clean environment.	3.15/.982	.820
		[MNC] seems to make an effort to create new jobs.	3.36/.861	.859
		[MNC] seems to be environmentally responsible.	3.40/.933	.897
	Reliable/financially strong company (RFSC)	[MNC] appears to have strong prospects for future growth.	3.73/.883	.907
		[MNC] seems to recognize and take advantage of market opportunities.	3.71/.875	.898
		[MNC] tends to outperform competitors.	3.64/.883	.874
Corporate reputation (parcels)		CO	3.52/.808	.870
		GE	3.58/.806	.902
		PRQ	3.79/.833	.904
		SER	3.30/.841	.796
		RFSC	3.69/.819	.880
Loyalty (Oliver 1999, 2015; Walsh and Beatty 2007)		I am a loyal customer of [MNC].	3.09/1.101	.730
		I have developed a good relationship with [MNC].	3.60/1.015	.913
		I am certain that I will buy products/offers of [MNC].	3.58/1.049	.930

Notes: MV/Std. = Mean value/standard deviation; λ = Standardized factor loadings (confirmatory factor analysis) (≥ .5).

Tab. 4: Measurement

and semantic equivalence was ensured (e.g., using the translation-back-translation method; Hult et al. 2008). The latter is continually discussed in international business research (e.g., Peterson et al. 2012). Although CR is conceptualized with 15 items in five dimensions, i.e., as a second-order construct (Walsh and Beatty 2007), we subsequently employed item parcelling to reduce the model complexity.[3]

On level 2, country measurements of EMB were based on Schwartz (1994) and obtained the most recent data from Shalom H. Schwartz at Hebron University. To measure CD, several indicators are discussed in the literature (e.g., GDP per capita, Chan et al. 2008; GNI per capita, Jin et al. 2015). We rely on one of the most frequently employed indicators – the Human Development Index (e.g., Zarantonello et al. 2013). Data for the respective year was obtained from the United Nations Development Programme (2017). Missing country data are frequently replaced with data from the nearest available neighbouring country (e.g., Walsh et al. 2014), including robustness checks. However, as this procedure is controversial in international business research (e.g., Cuervo-Cazurra et al. 2016), we have not used it.

The choice of covariates is not MSEM-specific but an advantage of MLM is that it facilitates the inclusion of covariates at any level (when theoretically or methodologically useful; e.g., Snijders and Bosker 2012, p. 56). For example, Zhou et al. (2010) controlled for consumer ethnocentrism and consumer bias on level 2 in favour of local brands for theoretical reasons. However, a restriction on the number of covariates is the number of level 2 clusters, i.e., many covariates may challenge the model identification. In our example, we controlled for age and gender (0 = male, 1 = female) on level 1 as they may influence CR perceptions and loyalty and the number of respondents per country on level 2 due to unequal numbers of respondents across countries.

Reliability and validity

For hierarchical data, tests for multilevel reliability are required, i.e., reliability indices have to be computed in an MLM-based manner because the common single-level reliability estimates do not reflect the true scale reliability at any specific level of analysis. Because the traditional, single-level reliability is the weighted average of between and within reliabilities, the overall reliability can be satisfactory when one of the level-specific reliabilities is poor. Recent statistical procedures enabled the computation of multilevel reliability measures (e.g., Geldhof et al. 2014); however, in MSEM studies, traditional tests, e.g., item-to-total correlation, Cronbach's alpha, or composite reliability, are still applied (e.g., Swoboda et al. 2017 were not aware of these requirements). Three options for reliability tests have been proposed for MSEM (Geldhof et al. 2014): multilevel alpha (α) (Cronbach 1951), multilevel composite reliability (ω) (McDonald 1999), and multilevel maximal reliability (H) (Hancock and Mueller 2001; Raykov 1997):

$$\alpha = \frac{n^2 \bar{\sigma}_{ij}}{\sigma_X^2} \tag{5}$$

where n is the number of items included in the scale, $\bar{\sigma}_{ij}$ is the average inter-item covariance within a scale, and σ_X^2 is the variance of the scale score;

$$\omega = \frac{(\sum_{i=1}^{k} \lambda_i)^2}{(\sum_{i=1}^{k} \lambda_i)^2 + \sum_{i=1}^{k} \theta_{ii}} \tag{6}$$

where λ_i is the factor loading of item i onto a single common factor, and θ_{ii} is the unique variance of item i; and

$$H = \left(1 + \frac{1}{\sum_{i=1}^{k} \frac{\iota_i^2}{1-\iota_i^2}}\right)^{-1} \tag{7}$$

where ι_i^2 is the squared standardized factor loading of indicator i onto a single common factor (Geldhof et al. 2014). First, a multilevel CFA (MCFA) model is fit. Second, the calculated estimates are inserted into these standard formulas to separately calculate reliability estimates at each level (e.g., α_B; α_W). We recommend reporting multilevel α, ω, and H in future studies. In our literature review, only Swoboda and Hirschmann (2017) additionally tested for multilevel reliability.

Specific multilevel validity measures are not available. Construct validity needs to be tested using traditional CFA factor loadings, average variance extracted (AVE), and discriminant validity tests, including the comparison of level-specific AVE values with squared correlations (Fornell and Larcker 1981) or the recent Heterotrait-monotrait (HTMT) method (Voorhees et al. 2016).

In our example, the multilevel alpha, composite reliability, and maximal reliability (see *Tab. 5*), as well as the construct validity, AVE, and discriminant validity (see *Tab. 6*–7) were satisfactory. We correlated the variables (see *Tab. 8*) (correlations under 0.80 are assumed to be acceptable, Zhou et al. 2010; however, this is a seperate issue not addressed in this paper).

	Alpha		Composite reliability		Maximal reliability	
	α_W	α_B	ω_W	ω_B	H_W	H_B
CO	.915	.994	.916	.996	.919	.998
GE	.918	.994	.918	.995	.919	.998
PRQ	.911	.994	.912	.994	.914	.996
SER	.879	.985	.877	.985	.884	.993
RFSC	.906	.995	.907	.995	.908	.998
Loy	.866	.975	.868	.981	.909	1.000
CR (parcels)	.928	.989	.928	.990	.934	.996

Notes: CO = Customer orientation, GE = Good employer, PRQ = Product range quality, SER = Social and environmental responsibility, RFSC = Reliable and financial strong company, LOY = Loyalty, α = Alpha (≥ .8), ω = composite reliability (≥ .8), H = maximal reliability (≥ .8).

Tab. 5: Multilevel reliability before and after item parcelling

	CO	GE	PRQ	SER	RFSC	LOY
CO	**.873**	*.762*	*.679*	*.651*	*.635*	*.564*
GE	.873 ***	**.879**	*.750*	*.643*	*.728*	*.593*
PRQ	.824 ***	.866 ***	**.862**	*.582*	*.792*	*.681*
SER	.807 ***	.802 ***	.763 ***	**.811**	*.573*	*.593*
RFSC	.797 ***	.853 ***	.890 ***	.757 ***	**.860**	*.608*
LOY	.751 ***	.770 ***	.825 ***	.770 ***	.780 ***	**.810**

Confirmatory model fit of single CR dimensions: CFI .979; TLI .973; RMSEA .048; SRMR .026; θ²(120) = 9524.429; scaling correction factor mean-adjusted maximum likelihood = 1.3975.
Notes: CO = Customer orientation, GE = Good employer, PRQ = Product range quality, SER = Social and environmental responsibility, RFSC = Reliable and financial strong company, LOY = Loyalty; AVE = Average variance extracted (≥ .5); AVEs are on the diagonal; squared correlations are above the diagonal; correlations are below the diagonal.
* $p < .05$; ** $p < .01$; *** $p < .001$; ns = not significant.

Tab. 6: Discriminant validity before item parcelling

	CR	LOY
CR	**.863**	*.731*
LOY	.855 ***	**.811**

Confirmatory model fit of parcelled CR dimensions: CFI .968; TLI .953; RMSEA .092; SRMR .026; θ²(19) = 5514.402; scaling correction factor mean-adjusted maximum likelihood = 1.3543.
Notes: CR = Corporate reputation, LOY = Loyalty; AVE = Average variance extracted (≥ .5); AVEs are on the diagonal; squared correlations are above the diagonal; correlations are below the diagonal.
* $p < .05$; ** $p < .01$; *** $p < .001$; ns = not significant.

Tab. 7: Discriminant validity after item parcelling

		1.	2.	3.	4.	5.	6.	7.
1.	CR	1						
2.	LOY	.855***	1					
3.	Gender	.009ns	.017*	1				
4.	Age	-.064***	-.093***	-.006ns	1			
5.	Group size	–	–	–	–	1		
6.	EMB	–	–	–	–	-.149***	1	
7.	CD	–	–	–	–	.152***	-.754***	1

Note: CR = Corporate reputation, LOY = Loyalty, EMB = Embeddedness, CD = Country development.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns = not significant.

Tab. 8: Correlations

Measurement invariance

MSEM requires testing for multilevel MI, which ensures that the modelled constructs equally measure the targeted constructs across clusters, e.g., whether respondents with different national cultures interpret a measure in a conceptually similar manner. Traditional methods (e.g., multi-group factor analysis, which tests the equality of measurement parameters over groups) are employed but are infeasible with a large number of groups (Selig et al. 2008) and justify the functioning of constructs at one level, which are used to investigate relationships at multiple levels (Zyphur et al. 2008). For example, Swoboda et al. (2016) applied multi-group CFA. In contrast, multilevel MI randomly treats group membership and tests for violations of MI across clusters (cluster bias). Some options were suggested (e.g., Jak et al. 2013; Muthén and Asparouhov 2013; in general, see Zyphur et al. 2008). We recommend using the procedure by Jak et al. (2013) as it is meaningful. The procedure is three-fold: ICCs, i.e., computation of a null and an independent model are required to confirm the necessity of multilevel analysis of the nested data structure, followed by a measurement model and a test for cluster bias. Jak et al. (2013) recommended the use of robust maximum likelihood (MLR) estimation for all computations because it offers a test statistic that is asymptotically equivalent to the T2 test statistic of Yuan and Bentler (2000). RMSEA is used as an approximate fit index: level-specific RMSEAs ($RMSEA_W$ for the within-level and $RMSEA_B$ for the between-level) are computed (Ryu and West 2009).

In our example, all factor loadings were considered equal across levels (see *Tab. 9*a-c). For the LOY1-item, 2.72 % of the total variance is explained by cluster bias. MI is not a serious problem in our example.

Common method variance, endogeneity, and unobserved heterogeneity

No specific tests for common method variance (CMV) and endogeneity or unobserved heterogeneity in MLM exist; however, tests that incorporate multilevel data are reasonable.

Three options for avoiding *CMV* are common. First, CMV can be addressed a priori via an appropriate questionnaire design (e.g., variation in scale formats and avoidance of repeated use of same anchor points to re-

	Cross-level effect model	Cross-level (effect and) interaction model
ICCs		
CO	.133	–
GE	.170	–
PRQ	.161	–
SER	.147	–
RFSC	.177	–
LOY1	–	.142
LOY2	–	.191
LOY3	–	.208
Null model		
χ^2(df)	8977.842(15)	9217.633(6)
p	***	***
RMSEA	.132	.211
$RMSEA_B$	3.423	5.487
Independence model		
χ^2(df)	482.291(10)	336.911(3)
p	***	***
RMSEA	.037	.057
$RMSEA_B$	.962	1.040

Notes: CO = Customer orientation, GE = Good employer, PRQ = Product range quality, SER = Social and environmental responsibility, RFSC = Reliable and financial strong company.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns = not significant.

Tab. 9a: ICCs, null model, and independent model

	Cross-level effect model	Cross-level (effect and) interaction model
One-factor model		
χ^2(df)	656.637(5)	2324.404(19)
p	***	***
RMSEA	.062	.059
$RMSEA_W$	.062	.059
Residual covariance model 1[1]		
χ^2(df)	98.996(4)	1727.337(18)
p	***	***
RMSEA	.026	.053
$RMSEA_W$	.026	.053
Residual covariance model 2[2]		
χ^2(df)	–	1546.001(17)
p	–	***
RMSEA	–	.051
$RMSEA_W$	–	.051
Residual covariance model 3[3]		
χ^2(df)	–	1315.830(16)
p	–	***
RMSEA	–	.049
$RMSEA_W$	–	.049

Notes: [1]A residual covariance was added between PRQ and RFSC. [2]A residual covariance was added between PRQ and SER. [3]A residual covariance was added between CO and GE. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns = not significant.

Tab. 9b: Measurement model

	Cross-level effect model	Cross-level (effect and) interaction model
Cluster invariance model		
χ^2(df)	2294.948(17)	11469.934(40)
p	***	***
RMSEA	.062	.091
Cluster bias model[1]		
χ^2(df)	941.004(16)	3163.868(39)
p	***	***
RMSEA	.041	.048
Total variance explained by cluster bias	2.40%	2.72%

Notes: [1]The residual variance for SER was freed for the cross-level effect models, and the residual variance for LOY1 was freed for the cross-level (effect and) interaction models. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns = not significant.

Tab. 9c: Cluster bias

duce artificial covariation). Second, the weak single-factor test can be computed (Podsakoff et al. 2003). Third, the marker variable technique (Lindell and Whitney 2001), which is based on the latent variable approach, is established (Williams et al. 2010). Almost all MSEM studies in our literature review did not test for CMV (e.g., Kiersch and Byrne 2015).

The likelihood of possible *endogeneity* or *unobserved heterogeneity* should be considered because predictor variable estimates can become biased and inconsistent if important variables that are correlated with the predictor variable are omitted from a model (Wooldridge 2010, p. 129). Traditional endogeneity tests to reduce possible biases from omitted variables (Antonakis et al. 2014) suggest the selection of a theoretically related instrumental variable (IV) for each independent variable and calculation of the strength of IV using an F-test and an efficient model (Stock and Watson 2011). An insignificant difference between the efficient and consistent model indicates the heterogeneity of the independent variable (Hausman 1978). Additionally, computing a rival model with the reverse model configuration is possible; a proposed model is supported if the rival model's fit is significantly poorer. Options to address unobserved heterogeneity involve generalizing the traditional approaches to SEM and include finite-mixture modelling following the procedures by Jedidi et al. (1997), Becker et al. (2013) or Raykov et al. (2016) or additional genetic algorithm or reweight regression approaches. Almost all MSEM studies in our review did not test for endogeneity or unobserved heterogeneity.

In our cross-level interaction model-example, we applied all three options regarding *CMV*. Within the single-factor test, the model with all items loading on a single factor (CFI .895; TLI .853; RMSEA .162; SRMR .044; $\chi^2(20) = 18159.224$) exhibited significantly worse fit values than our model did ($\Delta\chi^2(1) = 12644.822$, $p < .001$). We used the available job variable as a marker because it is theoretically unrelated to our constructs. The tests did not reveal any significant changes in coefficients or cor-

Model	θ²	df	CFI	TLI	RMSEA	SRMR
CFA	5875.158	25	.969	.955	.082	.024
Baseline	6122.136	28	.967	.958	.080	.036
Method-C	5953.323	27	.968	.958	.080	.025
Method-U	6837.812	20	.963	.934	.100	.024
Method-R	6766.926	21	.964	.938	.097	.024
	Chi-square differences of model comparison tests					
ΔModels	Δθ²	Δdf	p			
Baseline with Method-C	168.813	1	***			
Method-C with Method-U	884.489	7	***			
Method-U with Method-R	70.886	1	ns			

Notes: * $p < .05$; ** $p < .01$; *** $p < .001$; ns = not significant.

Tab. 10a: Results of model comparisons (phase I)

	Reliability baseline model	Decomposed reliability from method-U model		
Latent variable	Total reliability	Substantive reliability	Method reliability	% reliability marker variable
CR	.940	.936	.004	4.26%
LOY	.896	.885	.010	1.12%
Job (Marker variable)	.399			

Notes: CR = Corporate reputation; LOY = Loyalty.

Tab. 10b: Results of the reliability decomposition (phase II)

Construct correlations	CFA	Baseline	Method-U	Method-S (.05)	Method-S (.01)
CR with LOY	.855	.855	.852	.838	.838
Job with CR	-.166	.000	.000	.000	.000
Job with LOY	-.190	.000	.000	.000	.000

Notes: CR = Corporate reputation; LOY = Loyalty.

Tab. 10c: Results of the sensitivity analyses (phase III)

relations (see *Tab. 10*a-c). Regarding *endogeneity*, we selected adapted offers; well organized, brand quality, environmental causes; and brand strength as IVs for each of the five CR dimensions (see Swoboda et al. 2016) because CR was originally conceptualized as a second-order construct. The F-values exceeded the recommended threshold of 10 (Antonakis et al. 2014), and the efficient model did not significantly differ from the consistent model (all z-values < 1.96). A rival model's fit – specifying consumer loyalty as an antecedent of CR – was significantly poorer ($\Delta\chi^2(9) = 746.556$, $p < .001$). Endogeneity did not appear to be a problem in this example. Additional unobserved heterogeneity tests are not presented in this paper (see, e.g., Swoboda et al. 2016).

4.3. Method

Software

For MLM, various software packages are available (see *Tab. 11*). These software packages cover certain considered models but differ in terms of other criteria (e.g., number of possible levels and ability to treat latent constructs). The vast majority of studies applied HLM, followed by MLWinN. Of the MSEM studies in our literature review, 19 studies applied Mplus. One study employed LatentGold or EQS.

In our example, we used Mplus 7.4, which treats the calculated model types (cross-level effect, cross-level interaction, and cross-level effect and interaction) and more complex models, such as MSEM-moderated mediation (e.g., Preacher et al. 2011).

Centring

In MLM, independent variables are often centred on a specific value. Although centring is not a requirement in MSEM, it is recommended to facilitate the meaning of the intercept at level 2. Two options of centring independent variables are common: *grand mean centring* (GMC) and *group mean centring/centring within clusters* (CWC) (e.g., Algina and Swaminathan 2011, p. 285; Raudenbush and Bryk 2002, p. 33). In GCM, the grand mean of the independent variable (here CR) is subtracted from each independent variable observation:

$$CR_{ij} - \bar{x}_{CR} \quad (8)$$

where x_{CR} is the average of all CR scores in the sample. In CWC, for each level 1 independent variable observation, the level 2 cluster mean of the independent variable is subtracted from the independent variable's observation:

$$CR_{ij} - \bar{x}_{CR_j} \quad (9)$$

	HLM	MLwiN	Stata	SAS	Mplus	R
Data						
Binary data	x	x	x	x	x	x
Ordinal data	x	x	x	x	x	x
Multinomial data	x	x	x	x	x	x
Dichotomous data	x	—	x	x	x	x
Count data	x	x	x	x	x	x
Censored data	—	—	x	x	x	x
Model						
Number of levels	4	5	3	3	3	2
Means as outcome	x	—	—	x	x	x
Random intercept model	x	x	x	x	x	x
Random slope model	x	x	x	x	x	x
Cross-level interaction	x	x	x	x	x	x
Cross-classification	x	x	x	x	x	x
Latent constructs	—	—	x	x	x	—
Longitudinal models	x	x	x	x	x	x

Notes: x available; — not available. Software versions: HLM 7, MLwiN 3.00, Stata 14, SAS 14.2, Mplus 7.4, R 2.6.

Tab. 11: Comparison of select software for MLM

where x_{CR_j} is the average of all CR observations in the analysis that belong to cluster *j*.

GMC facilitates model computation and the interpretation of interaction effects. With CWC, all information about the group means is removed from the model; therefore, it should be used with caution (Hox 2013, p. 290). Enders and Tofighi (2007) demonstrated that CWC should be employed when the focus is on a level 1 predictor as CGM eliminates all between-cluster variation from the predictor and produces 'pure' estimate of the pooled within-cluster coefficient. CGM should be used when a level 2 predictor is of primary interest. In cases in which the focus is on examining a predictor's influence at two levels, either form of centring is appropriate. Although most MLM studies report the type of centring, some do not (e.g., Minola et al. 2016). In our example, we use GMC.

Model specification

The model specifications are shown for three types of MSEM models, where we used the equations in our empirical example for illustration purposes. CR, LOY, and EMB (shown) can be replaced by *x*, *y*, and any level 2 moderator to obtain general equations.

To test a *cross-level effect*, we follow a stepwise procedure. In the first step, a model with no explanatory variables is analysed ("Null model"). The model only contains the explained variable to partition its variance into level 1 and level 2 variance and is used to calculate the ICC values (see above). In the second step, a baseline model that additionally contains control variables on levels 1 and 2 is defined. In the third step, means as outcomes models that explain the mean value differences in the dependent variable CR on level 1 through level 2 variables are computed. The level 1 equation for CR is expressed as follows:

$$CR_{ij} = \beta_{0j} + \beta_{controls} Controls_{ij} + r_{ij} \tag{10}$$

On level 1, a decomposition of CR into the country average (β_{0j}) and individual deviations from this average (r_{ij}) is performed, where *i* denotes consumers, *j* denotes countries, CR_{ij} reflects consumer *i*'s perception of the MNC's CR in country *j*, $Controls_{ij}$ includes level 1 control variables, and r_{ij} denotes the level 1 error term. On level 2, differences in the countries' CR means are explained by EMB. The level 2 equation is expressed as

$$\beta_{0j} = \gamma_{00} + \gamma_{01} (CCV_j) + u_{0j} \tag{11}$$

where CCV_j represents the EMB variable on level 2, and u_{0j} is the error term, i.e., the part of the countries' CR mean β_{0j} that cannot be explained by EMB. For each predictor *j*, the computation of a separate multilevel model for hypothesis testing is recommended.

Similar steps are recommended for testing *cross-level interactions* (e.g., Hox 2010; Raudenbush and Bryk 2002). In the first step, the "null model" is analysed. In the second step, a "random intercept baseline model", which includes level 1 control variables, is defined. The "random intercept full model" in the third step additionally contains the level 1 predictor CR. In the fourth step, a "random intercept and slope baseline model" employs a random intercept and slope, with additional level 2 control variables. A cross-level interaction between the explanatory level 2 variable EMB and the slope of the level 1 effect between the level 1 independent variable CR and the level 1 dependent variable LOY is added. This final model represents the following level 1 equation:

$$LOY_{ij} = \beta_{0j} + \beta_{1j}(CR_{ij}) + \beta_{controls} Controls_{ij} + r_{ij} \tag{12}$$

where *i* denotes consumers, *j* denotes countries, LOY_{ij} reflects consumer *i*'s loyalty in country *j*, CR_{ij} reflects consumer *i*'s perception of the MNC's CR in country *j*, and $Controls_{ij}$ denotes level 1 control variables. The intercept β_{0j} and the slope β_{1j} are allowed to vary across level 2 clusters. Finally, r_{ij} represents the level 1 error term. The level 2 model captures the differences between level 2 clusters and predicts the variation in the β coefficients using the level 2 variable EMB as a predictor. This model is specified as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\,(CCV_j) + u_{0j} \tag{13}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\,(CCV_j) + u_{1j} \tag{14}$$

where CCV_j denotes the different level 2 EMB, and u_{qj} (q = 0, 1) are errors that are normally distributed over respondents. The complete model comprises equations (12) to (14) and was employed for hypothesis testing in the effect models:

$$LOY_{ij} = \gamma_{00} + \gamma_{01}\,(CCV_j) + \gamma_{10}(CR_{ij}) + \gamma_{11}(CCV_j)(CR_{ij}) + \gamma_{controls}\,Controls_{ij} + error \tag{15}$$

Testing for *cross-level effects and interactions simultaneously* combines equations from the two models. Steps one through four are the same as the steps in the *cross-level interaction* models. In the fifth step, a cross-level effect between the explanatory level 2 variable EMB and the level 1 independent variable and a cross-level interaction between the explanatory level 2 variable EMB and the slope of the level 1 effect between the level 1 independent variable CR and the level 1-dependent variable LOY were added. The level 1 equations are expressed as

$$CR_{ij} = \beta_{0j} + r_{CRij} \tag{16}$$

$$LOY_{ij} = \beta_{1j} + \beta_{2j}(CR_{ij}) + \beta_{controls}Controls_{ij} + r_{LOYij} \tag{17}$$

where i denotes consumers, j denotes countries, CR_{ij} reflects consumer i's perception of the MNC's CR in country j, LOY_{ij} reflects consumer i's loyalty in country j, and $Controls_{ij}$ denotes level 1 control variables. The intercept β_{0j} and the slope β_{1j} can vary across level 2 clusters; r_{ij} represents the level 1 error term.

The level 2 equations are as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\,(CCV_j) + u_{0j} \tag{18}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\,(CCV_j) + u_{1j} \tag{19}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\,(CCV_j) + u_{2j} \tag{20}$$

where CCV_j denotes the different level 2 variables, and u_{qj} (q = 0, 1) are errors that are normally distributed over respondents. The used model comprises equations (16) to (20) and includes all effects:

$$\begin{aligned} LOY_{ij} = {} & \gamma_{10} + \gamma_{11}(CCV_j) + u_{1j} + (\gamma_{20} + \gamma_{21}(CCV_j) + u_{2j}) \\ & * (\gamma_{00} + \gamma_{01}(CCV_j) + u_{0j} + r_{CRij}) \\ & + \beta_{controls}Controls_{ij} + r_{LOYij} \end{aligned} \tag{21}$$

To run a cross-level interaction model in Mplus, the number of random sets of starting values need to be chosen, e.g., 1,000. The computational time can be as large as 10 (0.5) hours, and model convergence is not always guaranteed (see Muthén and Muthén 1998–2015).

4.4. Results, implications, and limitations

Scholars recommend presenting unstandardized coefficients in the results because standardized coefficients are not computable in random slope and intercept models with cross-level interactions (Raudenbush and Bryk 2002, p. 159). We explicitly note that unstandardized coefficients are typically small, especially for cross-level interactions, and standardized coefficients reported for the perception models (see *Tab. 12*–14) can differ from unstandardized coefficients due to standard deviations (Hox 2010, p. 305). A calculation of effect sizes might be additionally applied (following Marsh et al. 2009), which is computed as $(2*b*SD_{predictor})/SD_{outcome}$, where b is the unstandardized regression coefficient, $SD_{predictor}$ is the predictor's standard deviation, and $SD_{outcome}$ is the outcome variable's standard deviation. The effect size is comparable to Cohen's d (Cohen 1988) and underlines our reasoning.

Explained variance

Computing the explained variance is not a requirement for MSEM, but it tests research propositions regarding level-specific explained variances, e.g., strength of moderation. The concept of explained variance has been transferred from multiple linear regression to MLM by treating proportional reductions in the estimated variance components as analogues of R^2 (Snijders and Bosker 2012, p. 109). A straightforward approach to assess the proportion of explained variance is to examine the residual error variances in a sequence of models, e.g., null model vs. full model (Hox 2010, p. 70). The level 1 explained variance R_1^2 is calculated as

$$R_1^2 = \frac{(\sigma^2_{null} - \sigma^2_{full})}{\sigma^2_{null}} \tag{22}$$

where σ^2_{null} is the level 1 residual variance of the null model, and σ^2_{full} is the level 1 residual variance of the respective full model. Similarly, the level 2 explained variance R_2^2 is calculated as

$$R_2^2 = \frac{(\tau_{00/null} - \tau_{00/full})}{\tau_{00/null}} \tag{23}$$

where $\tau_{00/null}$ is the level 2 residual variance of the null model, and $\tau_{00/full}$ is the level 2 residual variance of the respective full model (Raudenbush and Bryk 2002). A further advantage of this method is comparable to the logic of partitioning unexplained variance in two-level models. They are the only measures available that describe level-specific explained variance (LaHuis et al. 2014). Explained variance has frequently been interpreted in MSEM-studies. However, explained variance can be a challenge when it becomes negative (Hox 2013). LaHuis et al. (2014) discuss alternatives for this case.

Model fit indices

Fit indices determine how well the constraints implied in a proposed model conform to the actual data. Fit indices for MSEM differ from those for HLM. HLM-based studies report adjusted R^2, acc. to OLS regression (e.g., Swoboda and Pennemann 2014), or the general -2 Log-likelihood (e.g., Ahearne et al. 2013). MSEM requires reporting known indices from SEM.

Several options for reporting model fit in MSEM exist. Typical structural fit indices are the model chi^2, approxi-

		Null model	Baseline model	Means as outcomes model	
			b *p*	b *p*	b *p*
Main effects					
EMB	→ CR			.479*** (.554)	
CD	→ CR				-.020*** (-.751)
Covariates individual level					
Gender	→ CR		.001ns	.001ns	.001ns
Age	→ CR		.031***	.031***	.031***
Covariates country level					
Group size	→ CR		.000ns	.000ns	.000*
Residual variance (individual level)		.418	.416	.416	.416
Residual variance (country level)		.084	.090	.065	.046
Explained variance (individual level)			.5%		
Explained variance (country level)				27.8%	48.9%
AIC		259116.90	259020.39	259006.08	258988.44
BIC (adjusted)		259328.07	259256.89	259251.03	259233.39

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns = not significant.

Notes: Standardized values are shown in brackets; CR = Corporate reputation, EMB = Embeddedness, CD = Country development.

Tab. 12: Results of cross-level effects

mate fit indices (CFI, RMSEA, SRMR; with common cut-off criteria in SEM), and fit indices for model comparison (AIC, BIC; models in which the smaller value is preferred). Hox (2010, p. 307) suggested reporting several fit indices, which MSEM studies often do not do. For example, Zhou et al. (2010) or Kiersch and Byrne (2015) only reported fit indices for their preliminary CFA analyses instead of the complete MSEM. However, in MSEM, fit indices are challenging as they apply to the entire model and are not level-specific. Because all fit indices depend to some extent on the sample size and the sample size for the between model is almost smaller than the sample size for the within model, the latter dominates the fit index value. Thus, whether the model fits (poorly) at level 1, level 2, or both levels is unclear. Ryu (2014) discussed options to overcome these challenges. The first option is a two-step procedure that computes single-level fit statistics at each level based on estimates of saturated covariance matrices; the second option computes partially saturated models. Simulation studies reveal that both options better detected misfit than the level-neglecting standard approach. Unfortunately, both options are currently not part of the available MSEM software.

In our example, we computed AIC and BIC (fit indices for model comparison). They are the commonly reported fit indices in MLM and provided by Mplus. Additionally, stability tests may be drawn, e.g., using data for the MNC from another year or for further MNCs.

Results

The results in our example indicate a reinforcing role of EMB ($b = .479$, $p < 0.001$) and a diminishing role of CD ($b = -.020$, $p < 0.001$) (see *Tab. 12*). CR perceptions are stronger (weaker) in countries that score high in terms of EMB (CD). CD explains a higher level of country-level variance (48.9 %).

Regarding CR-loyalty relationships (see *Tab. 13*), the results support the elsewhere hypothesized effects of EMB ($b = 1.410$, $p < 0.01$) and CD ($b = -.009$, $p < 0.001$). CR effects are reinforced in countries with high EMB (similar Swoboda et al. 2016) and diminished in countries with high CD. EMB and CD explain high levels (25.0 % and 40.0 %) of country-level variance.

In the models with simultaneous *cross-level effects and interactions* – which we present for the first time in the literature – the results are similar (see *Tab. 14*): EMB reinforces CR perceptions ($b = .471$, $p < 0.001$) and CR effects ($b = .880$, $p < 0.001$), and CD diminishes CR perceptions ($b = -.019$, $p < 0.001$) and weakens CR effects ($b = -.013$, $p < 0.001$). EMB and CD explain high levels (32.7 % and 49.3 %) of country-level variance.

Brief implications and limitations

Our example primarily contributes to the understanding of how to design an MSEM study. Therefore, we cautiously provide a brief overview of the research implications and limitations, as our study uses extant data and covers only a single MNC in one industry. However, the example introduces the novel idea that both EMB and CD affect CR perceptions and the effects of an MNC across countries.

Concerning *research implications*, *CR perceptions* across nations are important indicators of MNCs' positioning differences, which are strongly linked to the negative effect of CD (explaining 48.9 % of the country-level variance). Concerning *CR effects* on loyalty, our results support extant research (e.g., Swoboda et al. 2016)

		Random intercept			Random intercept and slope		
		Null model	Baseline model	Full model	Baseline model	Cross-level interactions	
			b *p*	b *p*	b *p*	b *p*[1]	b *p*[1]
Main relationship individual level							
CR	→ Loy			.922***	.907***	.907***	.904***
Cross-level interactions							
EMB	→ Intercept Loy					.240***	
	→ Slope CR					1.410**	
CD	→ Intercept Loy						-.009***
	→ Slope CR						-.009***
Covariates individual level							
Gender	→ Loy		.016ns	.015*	.015*	.015*	.015*
Age	→ Loy		.023**	-.004ns	-.003ns	-.003ns	-.003ns
Covariates country level							
Group size	→ Loy				.000ns	.000***	.000**
Residual variance (individual level)		.505	.504	.166	.158	.158	.158
Residual variance (country level)		.153	.159	.021	.020	.015	.012
Explained variance (individual level)			.2%	67.1%	4.8%		
Explained variance (country level)					4.8%	25.0%	40.0%
AIC		227156.90	227109.36	467429.74	469002.90	469021.764	469363.81
BIC (adjusted)		227266.70	227236.06	467700.04	469256.30	469317.397	469659.45

*p < 0.05; **p < 0.01; ***p < 0.001; ns = not significant.

Notes: Effect sizes are shown in brackets; CR = Corporate reputation, Loy = Loyalty, EMB = Embeddedness, CD = Country development; [1]Slopes used for hypothesis tests.

Tab. 13: Results of cross-level interactions

because EMB provides a strong explanatory power (25.0 % of the variance). CD is negative and explains 40.0 %. We conclude an important role of EMB (as a moderator) and a stronger role of CD (as antecedence and moderator). Future research may particularly use CD (not only culture) to explain diminishing CR perceptions and effects across nations.

We suggest addressing three *limitations*. First, concerning the data, an analysis of additional countries or industries, a broader set of MNCs, or different stakeholder groups would be advantageous (although they would also create methodological challenges, i.e., additional levels). We also use a homogenous consumer sample with some advantages/disadvantages. Second, alternative measures for CR or loyalty may change the results, and alternative measures of national variables are obvious. Third, future studies may extend the proposed conceptual model, for example, by analysing the five perceived CR dimensions as antecedents of loyalty across nations to determine which CR-dimension – e.g., quality or responsibility – most affects behaviour.

5. Directions and challenges for future research

Because recent developments in statistical theory have increased the MLM options, whereas MSEM is still seldom used, we attempted to provide a systematic approach for the use of the latter. Our first objective was to introduce scholars to MSEM and its advantages relative to HLM and to provide novel insights into the topics and limitations of MSEM/HLM use in extant studies. Our second objective was to systematically address the requirements, options, and challenges of MSEM in sampling, measurement, and method by referring to the literature and by providing an example.

Our literature review underlines that MSEM is seldom used in marketing and management research, whereas in marketing, MSEM is only used in international marketing research. Few studies discuss MSEM, though latent constructs prevail and multilevel settings are common (e.g., concerning global brands and customer segments). The main advantage of MSEM should be prevalent: it provides new theoretical insights and propositions when combining constructs and variables from different levels of data by considering cross-level interactions. Because MNCs often manage their international activities across nations, similar options exist for international management research, where manifest variables are more common (e.g., Peterson et al. 2012). However, the advantages of MSEM are relatively unexplored, particularly in general marketing. We mention areas of future research, e.g., promotion effects across time and sales teams' performance within organizations, perceived values of brands with various brand equities, firms' evaluation across sectors, or touchpoints in e-commerce.

We provided guidelines for the design of MSEM analyses. Based on our nontechnical explanation of the re-

		Random intercept				Random intercept and slope			
		Null model	Baseline model	Full model		Baseline model		Cross-level effect and interactions	
			b p	b p	b p	b p	b p	b p[1]	b p[1]
Main relationship individual level									
CR	→ Loy			.917***	.917***	.920***	.920***	.919***	.918***
Cross-level effects									
EMB	→ CR			.458***		-.044ns		.471***	
CD	→ CR				.001ns		.000ns		-.019***
Cross-level interactions									
EMB	→ Intercept Loy							.642***	
	→ Slope CR							.880***	
CD	→ Intercept Loy								-.024***
	→ Slope CR								-.013***
Covariates individual level									
Gender	→ Loy			.015*	.015*	.015*	.015*	.015*	.015*
Age	→ Loy			-.003ns	-.003ns	-.003ns	-.003ns	-.003ns	-.003ns
Covariates country level									
Group size	→ Loy			.000ns	.000ns	.000ns	.000ns	.000*	.000ns
Residual variance CR	(individual level)	.412	.412	.412	.412	.395	.396	.396	.396
Residual variance Loy	(individual level)	.512	.504	.166	.166	.158	.158	.158	.158
Residual variance CR	(country level)	.103	.104	.109	.106	.104	.100	.072	.054
Residual variance Loy	(country level)	.156	.160	.155	.155	.150	.150	.101	.076
Explained variance CR	(individual level)		.0%	.0%	.0%	4.1%	3.9%		
Explained variance Loy	(individual level)		1.6%	67.1%	67.1%	4.8%	4.8%		
Explained variance CR	(country level)					4.6%	5.7%	30.8%	46.0%
Explained variance Loy	(country level)					3.2%	3.2%	32.7%	49.3%
AIC		457045.27	486108.47	457040.08	457041.73	462103.49	462104.45	462120.54	462460.87
BIC (adjusted)		457349.35	486420.99	457377.95	457379.60	462382.23	462383.19	462441.51	462781.85

*p < 0.05; **p < 0.01; ***p < 0.001; ns = not significant.

Notes: Effect sizes are shown in brackets; CR = Corporate reputation, Loy = Loyalty, EMB = Embeddedness, CD = Country development; [1]Slopes are used for hypothesis tests.

Tab. 14: Results of cross-level effects and interactions

quirements, options, and challenges, we summarize the major promising directions and challenges for future research and organize them into conceptualization issues, methodological shortcomings, and general advice.

Scholars need to consider the following conceptualization issues in MSEM-based studies:

- Of paramount importance is the sampling procedure, especially the size of level 2 variables; the sample size should be selected based on the requirements of MSEM before data collection.
- Outcome variables can technically be situated at any level, which enables different model conceptualizations than HLM.
- Based on our experience, control variables may substantially affect the amount of explained variance; unlike in OLS regression, where R^2 is less affected by additional controls, in MSEM, they should be chosen carefully and used parsimoniously.

The following methodological shortcomings need to be developed further:

- Traditional tests, e.g., normality, validity, common method variance, endogeneity, and unobserved heterogeneity, do not take the multilevel structure into account. To the best of our knowledge, no methods have been developed yet to do so.
- For certain aspects, methods have been developed that consider the multilevel structure; however, they have not been implemented into multilevel software, such as Mplus. For example, we hope that multilevel fit values (e.g., Ryu 2014) will be implemented into MSEM software and developed further according to the requirements of MSEM models.
- To the best of our knowledge, concrete methods (i.e., formulas) to determine the number of required clusters are not available. As this issue is critical, statistical research in this direction seems useful.

Scholars should consider the following advice when conducting MSEM-based studies:

- MSEM is only feasible with multi-item constructs. When these have not been conceptualized or collected, MSEM is not possible.

- Very long computation times are critical, e.g., up to ten hours for a cross-level interaction model, such as those presented here. In Mplus, multiple CPU cores can be employed (if supported by the specific analysis type), which can accelerate the computation.
- Obtaining convergence of the model is challenging. In Mplus, for example, different starting value sets or the number of iterations should be employed to facilitate model conversion (Muthén and Muthén 1998–2015). A higher number of random sets of starting values and a larger number of iterations might help the model to converge but significantly increase the computational time (in our experience, increasing the number of starting values by 50 % generates computation times of 15 hours, but this strongly depends on the sample). We often conceptualize a theoretical model and then test whether it is statistically practicable (e.g., Preacher et al. 2016).

We elaborate three model types – cross-level effects, cross-level interactions, and cross-level effects and interactions – but additional model types are also promising, including mediation models, cross-classifications, and longitudinal designs, as are analyses of additional levels (e.g., Preacher et al. 2011). We believe that the use of MSEM will increase in future research and hope that this article helps an increasing number of scholars apply MSEM.

Notes

[1] We do not address analyses that control for confounding effects at one level while testing relationships at another level, and longitudinal MSEM/HLM (where repeated measurements are nested within an individual) are not addressed. Longitudinal MLM is applied to large-scale panel survey data (Big Data), with a focus on patterns of change over time, and it does not require balanced data. For the advantages of longitudinal MLM, see, e.g., Hox (2010).

[2] Screening criteria to select respondents were used: quota sampling according to gender and age distribution based on information provided by national registration offices in each country, restricted to urban dwellers between the ages of 18 and 65 (55) years in developed (emerging) countries with above-average income and high levels of education or professional employment. The choice was made for various reasons, e.g., sample comparability across nations. Because we used panels, we need to control for data and panel quality, e.g., prevent bogus responses, instructional manipulation checks, and control for straight-lining or random clicking (e.g., Kaminska et al. 2010).

[3] We created a first-order construct – after ensuring the reliability and validity of every second-order factor – by using an inexact but common procedure (e.g., Swoboda et al. 2016): coarse factor scores (of indicators in each CR dimension, item parcelling) vs. refined factor scores (estimates taking correlation matrix and factor analysis coefficients into account; e.g., Brown 2015, p. 32; DiStefano et al. 2009). The advantages of our procedure include ease of computation, preservation of variation in original data, and stability across samples. We assured that each second-order factor had approximately equal factor loadings (otherwise items with relatively low vs. high loadings will have the same weight; DiStefano et al. 2009, p. 3) and computed validity coefficients for each of the factors (i.e., correlations between the factor score estimates and their respective factors; Grice 2001).
MSEM allows the modelling of second-order constructs with advantages such as retention of the original factor structure, whereas the model complexity increases. However, the challenges dominate, e.g., increased computation times, notably complicated convergence (if the number of level-2 clusters is small; e.g., Preacher et al. 2016), and model non-identification. The second order is seldom applied, and scholars in our review who measured latent factors and reduced them to one item typically did not specify the factor score method used (e.g., Kang et al. 2015 measured transformational leadership as a second-order construct but were unclear about how it was entered into the MSEM).

References

Ahearne, M., Lam, S. K., Hayati, B., & Kraus, F. (2013). Intrafunctional Competitive Intelligence and Sales Performance: A Social Network Perspective. *Journal of Marketing*, 77(5), 37–56.

Aitkin, M., & Longford, N. (1986). Statistical Modelling Issues in School Effectiveness Studies. *Journal of the Royal Statistical Society*, 149(1), 1–43.

Algina, J., & Swaminathan, H. (2011). Centering in Two-Level Nested Designs. In J. J. Hox, & J. K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis*, New York: Routledge, 285–312.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). Causality and Endogeneity: Problems and Solutions. In D. V. Day (Ed.), *The Oxford Handbook of Leadership and Organizations*, Oxford: Oxford University Press, 93–117.

Aryee, S., Walumbwa, F. O., Seidu, E. Y. M., & Otaye, L. E. (2016). Developing and Leveraging Human Capital Resource to Promote Service Quality: Testing a Theory of Performance. *Journal of Management*, 42(2), 480–499.

Atkins, D. C. (2005). Using Multilevel Models to Analyze Couple and Family Treatment Data: Basic and Advanced Issues. *Journal of Family Psychology*, 19(1), 98–110.

Bagozzi, R. P. (1977). Structural Equation Models in Experimental Research. *Journal of Marketing Research*, 209–226.

Balka, K., Raasch, C., & Herstatt, C. (2014). The Effect of Selective Openness on Value Creation in User Innovation Communities. *Journal of Product Innovation Management*, 31(2), 392–407.

Barile, J. P. (2016). Multilevel Structural Equation Modeling. In L. Jason, & D. Glenwick (Eds.), *Handbook of Methodological Approaches to Community-Based Research*, New York: Oxford University Press, 153–164.

Bartikowski, B., Walsh, G., & Beatty, S. E. (2011). Culture and Age as Moderators in the Corporate Reputation and Loyalty Relationship. *Journal of Business Research*, 64(9), 966–972.

Becker, J.-M., Rai, A., Ringle, C. M., & Völckner, F. (2013). Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats. *Mis Quarterly*, 37(3), 665–694.

Berry, H., Guillén, M. F., & Zhou, N. (2010). An Institutional Approach to Cross-National Distance. *Journal of International Business Studies*, 41(9), 1460–1480.

Blalock, H. M. (1984). Contextual-Effect Models: Theoretical and methodological issues. In R. H. Tuner, & J. F. Short Jr (Eds.), *Annual Review of Sociology*, Belmont, CA: Wadsworth Publishing, 353–372.

Breckler, S. J. (1990). Applications of Covariance Structure Modeling in Psychology: Cause for Concern? *Psychological Bulletin*, 107(2), 260–273.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. New York; London: The Guilford Press.

Burstein, L. (1980). The Role of Levels of Analysis in the Specification of Educational Effects. In R. Dreeben, & J. A. Thomas (Eds.), *Analysis of Educational Productivity: Issues in Microanalysis*, Cambridge, Mass.: Ballinger, 119–190.

Byrne, B. M. (2012). *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*. New York, NY: Routledge.

Calli, M. K., Weverbergh, M., & Franses, P. H. (2012). The Effectiveness of High-Frequency Direct-Response Commercials. *International Journal of Research in Marketing*, 29(1), 98–109.

Chan, C. M., Isobe, T., & Makino, S. (2008). Which Country Matters? Institutional Development and Foreign Affiliate Performance. *Strategic Management Journal*, 29(11), 1179–1205.

Cheung, M. W.-L., & Au, K. (2005). Applications of Multilevel Structural Equation Modeling to Cross-Cultural Research. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(4), 598–619.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16(3), 297–334.

Cuervo-Cazurra, A., Andersson, U., Brannen, M. Y., Nielsen, B. B., & Rebecca Reuber, A. (2016). From the Editors: Can I Trust Your Findings? Ruling Out Alternative Explanations in International Business Research. *Journal of International Business Studies*, 47(8), 881–897.

Dahling, J. J., Taylor, S. R., Chau, S. L., & Dwight, S. A. (2015). Does Coaching Matter? A Multilevel Model Linking Managerial Coaching Skill and Frequency to Sales Goal Attainment. *Personnel Psychology*, 69, 863–894.

Deephouse, D. L., Newburry, W., & Soleimani, A. (2016). The Effects of Institutional Development and National Culture on Cross-national Differences in Corporate Reputation. *Journal of World Business*, 51(3), 463–473.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research & Evaluation*, 14(20), 1–11.

du Toit, S. H. C., & Du Toit, M. (2008). Multilevel Structural Equation Modeling. In J. d. Leeuw, & E. Meijer (Eds.), *Handbook of Multilevel Analysis*, New York: Springer, 435–478.

Enders, C. K., & Tofighi, D. (2007). Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look At An Old Issue. *Psychological Methods*, 12(2), 121–138.

Falkenreck, C., & Wagner, R. (2010). Impact of Direct Marketing Activities on Company Reputation Transfer Success: Empirical Evidence from Five Different Cultures. *Corporate Reputation Review*, 13(1), 20–37.

Finch, W. H., & Bolin, J. E. (2017). *Multilevel Modeling Using Mplus*. Boca Raton: CRC Press.

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 39–50.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability Estimation In a Multilevel Confirmatory Factor Analysis Framework. *Psychological Methods*, 19(1), 72–91.

Gielnik, M. M., Spitzmuller, M., Schmitt, A., Klemann, D. K., & Frese, M. (2015). "I Put in Effort, Therefore I Am Passionate": Investigating the Path From Effort to Passion in Entrepreneurship. *Academy of Management Journal*, 58(4), 1012–1031.

Grice, J. W. (2001). Computing and Evaluating Factor Scores. *Psychological Methods*, 6(4), 430.

Haans, H., & Gijsbrechts, E. (2011). "One-Deal-Fits-All?" On Category Sales Promotion Effectiveness in Smaller versus Larger Supermarkets. *Journal of Retailing*, 87(4), 427–443.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking Construct Reliability Within Latent Variable Systems. In R. Cudeck, S. H. C. Du Toit, & D. Sörbom (Eds.), *Structural Equation Modeling, Present and Future*, Lincolnwood, IL: Scientific Software International, 195–216.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica: Journal of the Econometric Society*, 46(6), 1251–1271.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A regression-Based Approach*. New York, NY: Guilford Press.

Heck, R. H., & Thomas, S. L. (2015). *An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches Using Mplus*. New York, NY: Routledge.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing Measurement Invariance of Composites Using Partial Least Squares. *International Marketing Review*, 33(3), 405–431.

Hohenberg, S., & Homburg, C. (2016). Motivating Sales Reps for Innovation Selling in Different Cultures. *The Journal of Marketing*, 80(2), 101–120.

Homburg, C., Wieseke, J., & Hoyer, W. D. (2009). Social Identity and the Service-Profit Chain. *The Journal of Marketing*, 73(2), 38–54.

Hoppner, J. J., & Griffith, D. A. (2015). Looking Back to Move Forward: A Review of the Evolution of Research in International Marketing Channels. *Journal of Retailing*, 91(4), 610–626.

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications*. New York: Routledge.

Hox, J. J. (2013). Multilevel Regression and Multilevel Structural Equation Modeling. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods*, Oxford: Oxford University Press, 281–294.

Hult, G. T. M., Ketchen, 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.

Ieva, F., & Paganoni, A. M. (2015). Detecting and Visualizing Outliers in Provider Profiling via Funnel Plots and Mixed Effect Models. *Health Care Management Science*, 18(2), 166–172.

Jak, S., Oort, F. J., & Dolan, C. V. (2013). A Test for Cluster Bias: Detecting Violations of Measurement Invariance Across Clusters in Multilevel Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(2), 265–282.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity. *Marketing Science*, 16(1), 39–59.

Jensen, J. M., Patel, P. C., & Messersmith, J. G. (2013). High-Performance Work Systems and Job Control: Consequences for Anxiety, Role Overload, and Turnover Intentions. *Journal of Management*, 39(6), 1699–1724.

Jin, Z., Lynch, R., Attia, S., Chansarkar, B., Gülsoy, T., Lapoule, P., Liu, X., Newburry, W., Nooraini, M. S., & Parente, R. (2015). The Relationship Between Consumer Ethnocentrism, Cosmopolitanism and Product Country Image Among Younger Generation Consumers: The Moderating Role of Country Development Status. *International Business Review*, 24(3), 380–393.

Jöreskog, K. G. (1969). A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika*, 34(2), 183–202.

Jöreskog, K. G. (1971). Simultaneous Factor Analysis in Several Populations. *Psychometrika*, 36(4), 409–426.

Julian, M. W. (2001). The Consequences of Ignoring Multilevel Data Structures in Nonhierarchical Covariance Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), 325–352.

Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing Among Reluctant Respondents in a Cross-National Context. *Public Opinion Quarterly*, 74(5), 956–984.

Kang, J. H., Solomon, G. T., & Choi, D. Y. (2015). CEOs' Leadership Styles and Managers' Innovative Behaviour: Investigation of Intervening Effects in an Entrepreneurial Context. *Journal of Management Studies*, 52(4), 531–554.

Kelloway, E. K. (2015). *Using Mplus for Structural Equation Modeling: A Researcher's Guide*. Los Angeles: SAGE.

Kiersch, C. E., & Byrne, Z. S. (2015). Is Being Authentic Being Fair? Multilevel Examination of Authentic Leadership, Justice, and Employee Outcomes. *Journal of Leadership & Organizational Studies*, 22(3), 292–303.

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*. New York: Guilford Press.

Kreft, I. G. G., & Leeuw, J. d. (1998). *Introducing Multilevel Modeling*. London: SAGE.

LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained Variance Measures for Multilevel Models. *Organizational Research Methods*, 17(4), 433–451.

Laird, N. M., & Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, 963–974.

Larivière, B., Keiningham, T. L., Aksoy, L., Yalçin, A., Morgeson III, F. V., & Mithas, S. (2016). Modeling Heterogeneity in the Satisfaction, Loyalty Intention, and Shareholder Value Linkage: A Cross-Industry Analysis at the Customer and Firm Levels. *Journal of Marketing Research*, 53(1), 91–109.

LePine, M. A., Zhang, Y., Crawford, E. R., & Rich, B. L. (2016). Turning Their Pain to Gain: Charismatic Leader Influence on Follower Stress Appraisal and Job Performance. *Academy of Management Journal*, 59(3), 1036–1059.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for Common Method Variance in Cross-Sectional Research Designs. *Journal of applied psychology*, 86(1), 114.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The Multilevel Latent Covariate Model: A New, More Reliable Approach to Group-Level Effects in Contextual Studies. *Psychological Methods*, 13(3), 203–229.

Maas, C. J. M., & Hox, J. J. (2004). Robustness Issues in Multilevel Regression Analysis. *Statistica Neerlandica*, 58(2), 127–137.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology*, 1(3), 86–92.

Magnusson, P., Peterson, R., & A. Westjohn, S. (2014). The Influence of National Cultural Values on the Use of Rewards Alignment to Improve Sales Collaboration. *International Marketing Review*, 31(1), 30–50.

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-Latent Models of School Contextual Effects: Integrating Multilevel and Structural Equation Approaches to Control Measurement and Sampling Error. *Multivariate Behavioral Research*, 44(6), 764–802.

Martin, K. D., & Hill, R. P. (2015). Saving and Well-Being at the Base of the Pyramid: Implications for Transformative Financial Services Delivery. *Journal of Service Research*, 18(3), 405–421.

Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and Estimating the Power to Detect Cross-Level Interaction Effects in Multilevel Modeling. *Journal of applied psychology*, 97(5), 951.

McDonald, R. P. (1994). The Bilevel Reticular Action Model for Path Analysis with Latent Variables. *Sociological Methods & Research*, 22(3), 399–413.

McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Mehta, P. D., & Neale, M. C. (2005). People Are Variables Too: Multilevel Structural Equations Modeling. *Psychological Methods*, 10(3), 259–284.

Minola, T., Donina, D., & Meoli, M. (2016). Students Climbing the Entrepreneurial Ladder: Does University Internationalization Pay Off? *Small Business Economics*, 47(3), 565–587.

Mossholder, K. W., & Bedeian, A. G. (1983). Cross-Level Inference and Organizational Research: Perspectives on Interpretation and Application. *Academy of Management Review*, 8(4), 547–558.

Muthén, B., & Asparouhov, T. (2013). New Methods for the Study of Measurement Invariance with Many Groups.

Muthén, B. O. (1994). Multilevel Covariance Structure Analysis. *Sociological Methods & Research*, 22(3), 376–398.

Muthén, B. O. (2002). Beyond SEM: General Latent Variable Modeling. *Behaviormetrika*, 29(1), 81–117.

Muthén, B. O., & Muthén, L. K. (1998–2017). Mplus Discussion Forum.

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Newsom, J. T. (2002). A Multilevel Structural Equation Model for Dyadic Data. *Structural equation modeling*, 9(3), 431–447.

Ngobo, P. V., & Fouda, M. (2012). Is 'Good' Governance Good for Business? A Cross-National Analysis of Firms in African Countries. *Journal of World Business*, 47(3), 435–449.

Nielsen, B. B., & Nielsen, S. (2013). Top Management Team Nationality Diversity and Firm Performance: A Multilevel Study. *Strategic Management Journal*, 34(3), 373–382.

Oliver, P. H., Gottfried, A. W., Guerin, D. W., Gottfried, A. E., Reichard, R. J., & Riggio, R. E. (2011). Adolescent Family Environmental Antecedents to Transformational Leadership Potential: A Longitudinal Mediational Analysis. *The Leadership Quarterly*, 22(3), 535–544.

Oliver, R. L. (1999). Whence Consumer Loyalty? *Journal of Marketing*, 63(4), 33–43.

Oliver, R. L. (2015). *Satisfaction: A Behavioral Perspective On The Consumer*. New York, NY: Routledge.

Ozkaya, H. E., Dabas, C., Kolev, K., Hult, G. T. M., Dahlquist, S. H., & Manjeshwar, S. A. (2013). An Assessment of Hierarchical Linear Modeling in International Business, Management, and Marketing. *International Business Review*, 22(4), 663–677.

Peccei, R., & van de Voorde, K. (2016). The Application of the Multilevel Paradigm in Human Resource Management-Outcomes Research: Taking Stock and Going Forward. *Journal of Management*. 10.1177/0149206316673720

Peterson, M. F., Arregle, J.-L., & Martin, X. (2012). Multilevel Models in International Business Research. *Journal of International Business Studies*, 43(5), 451–457.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of applied psychology*, 88(5), 879–903.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative Methods for Assessing Mediation in Multilevel Data: The Advantages of Multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(2), 161–182.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel Structural Equation Models for Assessing Moderation Within and Across Levels of Analysis. *Psychological Methods*, 21(2), 189.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A General Multilevel SEM Framework for Assessing Multilevel Mediation. *Psychological Methods*, 15(3), 209.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized Multilevel Structural Equation Modeling. *Psychometrika*, 69(2), 167–190.

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2007). Multilevel Structural Equation Modeling. In S.-Y. Lee (Ed.), *Handbook of Latent Variable and Related Models*, Amsterdam; Boston: Elsevier/North-Holland, 209–227.

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2012). Multilevel Structural Equation Modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling*, New York: Guilford Press, 512–531.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CF: SAGE.

Raykov, T. (1997). Estimation of Composite Reliability for Congeneric Measures. *Applied Psychological Measurement*, 21(2), 173–184.

Raykov, T., Marcoulides, G. A., & Chang, C. (2016). Examining Population Heterogeneity in Finite Mixture Settings Using Latent Variable Modeling. *Structural equation modeling*, 23(5), 726–730.

Ryu, E. (2014). Model Fit Evaluation in Multilevel Structural Equation Models. *Frontiers in Psychology*, 5, 1–8.

Ryu, E., & West, S. G. (2009). Level-Specific Evaluation of Model Fit in Multilevel Structural Equation Modeling. *Structural equation modeling*, 16(4), 583–601.

Schwartz, S. H. (1994). Beyond Individualism-Collectivism: New Cultural Dimensions of Values. In U. Kim, H. C. Triandis, C. Kagitcibasi, & S. C. Choi (Eds.), *Individualism and Collectivism*, Thousand Oaks, CF: SAGE, 85–119.

Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent Variable Structural Equation Modeling in Cross-Cultural Research: Multigroup and Multilevel Approaches. In F. J. R. v. d. Vijver, D. A. van Hemert, & Y. H. Poortinga (Eds.), *Multilevel Analysis of Individuals and Cultures*, New York: Lawrence Erlbaum Associates, 93–119.

Shen, J., & Benson, J. (2016). When CSR is a Social Norm: How Socially Responsible Human Resource Management Affects Employee Work Behavior. *Journal of Management*, 42(6), 1723–1746.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Los Angeles: SAGE.

Spybrook, J., Raudenbush, S. W., Liu, X.-f., Congdon, R., & Martínez, A. (2011). Optimal Design for Longitudinal and Multilevel Research: Documentation for the "Optimal Design" Software.

Sridhar, S., & Srinivasan, R. (2012). Social Influence Effects in Online Product Ratings. *The Journal of Marketing*, 76(5), 70–88.

Stock, J. H., & Watson, M. W. (2011). *Introduction to Econometrics*. Boston: Addison-Wesley.

Swoboda, B., & Hirschmann, J. (2017). Perceptions and Effects of Cross-Natinoal Corporate Reputation: The Role of Hofstede's Cultural Value Approach. International Marketing Review.

Swoboda, B., Huber, C., Schuster, T., & Hirschmann, J. (2017). Corporate Reputation Effects Across Nations: The Impact of Country Distances and Firm-Specific Resources. *Management International Review*, DOI: 10.1007/s11575-017-0313–3.

Swoboda, B., & Pennemann, K. (2014). Do International Retailers benefit from being Global in Emerging Countries? A Multilevel Study in China. *Marketing ZFP – Journal of Research and Management*, 36(2), 141–150.

Swoboda, B., Puchert, C., & Morschett, D. (2016). Explaining the Differing Effects of Corporate Reputation Across Nations: A Multilevel Analysis. *Journal of the Academy of Marketing Science*, 44(4), 454–473.

United Nations Development Programme. (2017). Human Development Index. http://hdr.undp.org/en/data.

Venkatesan, R., & Farris, P. W. (2012). Measuring and Managing Returns from Retailer-Customized Coupon Campaigns. *The Journal of Marketing*, 76(1), 76–94.

Vernon, T., & Eysenck, S. (2007). Special Issue on Structural Equation Modeling. *Personality and Individual Differences*, 42(5).

Voorhees, C. M., Brady, M. K., Calantone, R., & Ramirez, E. (2016). Discriminant Validity Testing in Marketing: An Analysis, Causes for Concern, and Proposed Remedies. *Journal of the Academy of Marketing Science*, 44(1), 119–134.

Walsh, G., & Beatty, S. E. (2007). Customer-Based Corporate Reputation of a Service Firm: Scale Development and Validation. *Journal of the Academy of Marketing Science*, 35(1), 127–143.

Walsh, G., Beatty, S. E., & Shiu, E. M. K. (2009). The Customer-Based Corporate Reputation Scale: Replication and Short Form. *Journal of Business Research*, 62(10), 924–930.

Walsh, G., Shiu, E., & Hassan, L. M. (2014). Cross-National Advertising and Behavioral Intentions: A Multilevel Analysis. *Journal of International Marketing*, 22(1), 77–98.

Wieseke, J., Geigenmüller, A., & Kraus, F. (2012). On the Role of Empathy in Customer-Employee Interactions. *Journal of Service Research*, 15(3), 316–331.

Williams, L. J., Hartman, N., & Cavazotte, F. (2010). Method Variance and Marker Variables: A Review and Comprehensive CFA Marker Technique. *Organizational Research Methods*, 13(3), 477–514.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT Press.

Yi, Y., & Yoo, J. (2011). The Long-Term Effects of Sales Promotions on Brand Attitude Across Monetary and Non-Monetary Promotions. *Psychology & Marketing*, 28(9), 879–896.

Yuan, K. H., & Bentler, P. M. (2000). Three Likelihood-Based Methods for Mean and Covariance Structure Analysis with Nonnormal Missing Data. *Sociological Methodology*, 30(1), 165–200.

Zarantonello, L., Jedidi, K., & Schmitt, B. H. (2013). Functional and Experiential Routes to Persuasion: An Analysis of Advertising in Emerging Versus Developed Markets. *International Journal of Research in Marketing*, 30(1), 46–56.

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing Multilevel Mediation Using Hierarchical Linear Models: Problems and Solutions. *Organizational Research Methods*, 12(4), 695–719.

Zhou, L., Yang, Z., & Hui, M. K. (2010). Non-Local or Local Brands? A Multi-Level Investigation into Confidence in Brand Origin Identification and its Strategic Implications. *Journal of the Academy of Marketing Science*, 38(2), 202–218.

Zyphur, M. J., Kaplan, S. A., & Christian, M. S. (2008). Assumptions of Cross-Level Measurement and Structural Invariance in the Analysis of Multilevel Data: Problems and Solutions. *Group Dynamics: Theory, Research, and Practice*, 12(2), 127–140.

Keywords

Multilevel Structural Equation Modelling, MSEM; Hierarchical Linear Modelling, HLM, Multiple-level Modelling, Random Coefficients.