To cite this article:

Perez Alonso, A. F., Rosseel, Y., Vermunt, J. K., & De Roover, K. (in press). Mixture multigroup structural equation modeling: A novel method for comparing structural relations across many groups. *Psychological Methods*. https://doi.org/10.1037/met0000667

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

© 2024 The Author(s)
ISSN: 1082-989X

https://doi.org/10.1037/met0000667

Mixture Multigroup Structural Equation Modeling: A Novel Method for Comparing Structural Relations Across Many Groups

Andres F. Perez Alonso[1, 2], Yves Rosseel[3], Jeroen K. Vermunt[1], and Kim De Roover[1, 2]

[1] Department of Methodology and Statistics, Tilburg University

[2] Research Group of Quantitative Psychology and Individual Differences, KU Leuven

[3] Department of Data Analysis, Ghent University

Abstract

Behavioral scientists often examine the relations between two or more latent variables (e.g., how emotions relate to life satisfaction), and structural equation modeling (SEM) is the state-of-the-art for doing so. When comparing these "structural relations" among many groups, they likely differ across the groups. However, it is equally likely that some groups share the same relations so that clusters of groups emerge. Latent variables are measured indirectly by questionnaires and, for validly comparing their relations among groups, the measurement of the latent variables should be invariant across the groups (i.e., measurement invariance). However, across many groups, often at least some measurement parameters differ. Restricting these measurement parameters to be invariant, when they are not, causes the structural relations to be estimated incorrectly and invalidates their comparison. We propose mixture multigroup SEM (MMG-SEM) to gather groups with equivalent structural relations in clusters while accounting for the reality of measurement noninvariance. Specifically, MMG-SEM obtains a clustering of groups focused on the structural relations by making them cluster-specific, while capturing measurement noninvariances with group-specific measurement parameters. In this way, MMG-SEM ensures that the clustering is valid and unaffected by differences in measurement. This article proposes an estimation procedure built around the R package "lavaan" and evaluates MMG-SEM's performance through two simulation studies. The results demonstrate that MMG-SEM successfully recovers the group-clustering as well as the cluster-specific relations and the partially group-specific measurement parameters. To illustrate its empirical value, we apply MMG-SEM to cross-cultural data on the relations between experienced emotions and life satisfaction.

Translational Abstract

Behavioral scientists often examine the relations between two or more latent variables (e.g., unobservable variables such as attitudes or emotions), and structural equation modeling (SEM) is the state-of-the-art for doing so. When comparing these so-called "structural relations" among many groups (e.g., many countries), they likely differ across the groups. However, it is equally likely that some groups share the same relations and that clusters of groups emerge in terms of the relations. When comparing the latent variables' relations among groups, one should note that such variables are indirectly measured via questionnaire items, and one should evaluate whether this measurement is invariant across the groups (i.e., measurement invariance). In the case of many groups, at least some measurement parameters often differ across the groups. Existing mixture SEM methods capture these differences in measurement (which are

Douglas Steinley served as action editor.

Andres F. Perez Alonso https://orcid.org/0000-0002-2480-8771

Yves Rosseel https://orcid.org/0000-0002-4129-4477

Jeroen K. Vermunt https://orcid.org/0000-0001-9053-9330

Kim De Roover https://orcid.org/0000-0002-0299-0648

This research was funded by a Vidi Grant (VI.Vidi.201.133) awarded to Kim De Roover by the Netherlands Organization for Scientific Research (NWO). The code behind the simulations has been made publicly available on the GitHub. The work of this study was not preregistered. Ideas in this article were previously presented at the International Meeting of the Psychometrics Society 2022 and at the conference of the European Association of Methodology 2023.

This work is licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License (CC BY-NC-ND 4.0; https://creativecommons.org/licenses/by-nc-nd/4.0). This license permits copying and redistributing the work in any medium or format for noncommercial use provided the original authors and source are credited and a link to the license is included in attribution. No derivative works are permitted under this license.

Andres F. Perez Alonso served as lead for formal analysis, investigation, methodology, project administration, software, validation, visualization, and writing–original draft. Yves Rosseel served in a supporting role for conceptualization, investigation, methodology, software, supervision, and writing–review and editing. Jeroen K. Vermunt served in a supporting role for conceptualization, investigation, methodology, software, supervision, and writing–review and editing. Kim De Roover served as lead for funding acquisition, supervision, and writing–review and editing and served in a supporting role for formal analysis, investigation, methodology, project administration, software, validation, and visualization. Andres F. Perez Alonso and Kim De Roover contributed equally to conceptualization and resources.

Correspondence concerning this article should be addressed to Andres F. Perez Alonso, Department of Methodology and Statistics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: A.F.PerezAlonso@tilburguniversity.edu

not of interest) and the differences in the structural relations with the same clustering. We propose mixture multigroup SEM (MMG-SEM) to obtain a clustering of groups focused entirely on the structural relations by making them cluster-specific, while allowing for the measurement parameters to be (partially) group-specific. In this way, MMG-SEM disentangles differences in structural relations from differences in measurement parameters. In extensive simulation studies, MMG-SEM performed well in terms of recovering the clustering and the structural relations. Its empirical value is demonstrated for cross-cultural data on the relations between experienced emotions and life satisfaction.

Keywords: structural equation modeling, mixture modeling, multigroup modeling, structural relations, measurement invariance

Behavioral researchers often aim to compare the relationship between two or more constructs (e.g., attitudes, beliefs, and emotions) across many groups (e.g., countries and ethnic groups). For instance, Brandt et al. (2021) examined the association between people's feelings of threat and political beliefs across 56 countries, and Bastian et al. (2014) evaluated to what extent the social value of positive emotions determined life satisfaction across 47 countries. Structural equation modeling (SEM; Bollen, 1989; Hoyle, 2023) is the state-of-the-art for analyzing such relations, also called "structural relations," since it allows modeling larger path models—for instance, including mediation—with just one analysis (i.e., without performing several separate regression analyses).

To simultaneously fit a SEM model to several groups and compare their structural relations, multigroup SEM (MG-SEM) can be used. In the case of many groups, the structural relations will likely differ between the groups. For example, the link between feelings of threat and political ideology can vary between countries depending on their characteristics and historical background (Brandt et al., 2021). To gain insight into the differences and similarities across many groups, one must perform a daunting number of pairwise comparisons of the group-specific relations. For example, for 10 groups, it would be necessary to perform 45 pairwise comparisons, whereas, for the 56 groups by Brandt et al. (2021), it would entail no less than 1,540 pairwise comparisons.

Multilevel SEM is a popular parsimonious method for capturing differences in parameters among many groups (e.g., Davidov et al., 2012; Meuleman, 2019). Like other multilevel methods,[1] multilevel SEM deals with hierarchical data (e.g., people nested within countries) by adding random effects to capture differences among higher-level units in specific model parameters (e.g., structural relations; Varriale & Vermunt, 2012). This implies that each higher-level unit (e.g., country) has its own deviation from the overall parameter value, where these deviations are assumed to be normally distributed. Although the multilevel approach is a parsimonious approach in terms of parameter estimation, it is not at all parsimonious when it comes to comparing group-specific parameters such as structural relations. To pinpoint differences and similarities, one still needs to make the many pairwise comparisons of group-specific relations. Surely, in the multilevel approach, one can try to explain the differences by adding group-level characteristics—such as diversity and individualism—as moderators to the structural relations, but this implies inspecting a substantial amount of interaction effects that may still fail to explain the differences (e.g., Brandt et al., 2021).

An intuitive way to gain insight into the differences and similarities between the groups is to look for subsets or "clusters" of groups that have the same relations between constructs.[2] For instance, if countries with the same historical background share similar relations between feelings of threat and political beliefs, they will end up in the same cluster. In this article, we propose a new method that applies mixture modeling (McLachlan et al., 2019; McNicholas, 2017) to cluster groups based on their structural relations. This sounds straightforward, but an important complication is that structural relations are not the only source of differences that may be picked up by such clustering. In fact, the constructs of interest are latent variables, measured indirectly using a set of indicator variables, often questionnaire items. This implies that we need to consider the measurement model (MM), that is, which items measure which latent variable(s) and to what extent. SEM allows capturing the MM of latent variables (also called "factors" in SEM) and the structural relations between the latent variables (i.e., the structural model; SM) at the same time.

Before clustering the groups based on the factor's relations, one must ensure that the MM is similar enough across the groups to validly compare the relations. This assumption is called measurement invariance (MI; Meredith, 1993; Vandenberg & Lance, 2000). Specifically, equality of the so-called "factor loadings" should hold (i.e., the same items should measure a latent variable with the same reliability) for all groups, which is referred to as metric invariance. Indeed, since the factor loadings capture the item-factor relations, they should be invariant to enable valid comparisons of the relations among factors. If full metric invariance (i.e., equality of all loadings) does not hold, partial metric invariance (Byrne et al., 1989) may hold, which allows some of the loadings to be different or "noninvariant" across groups.[3] For example, Bastian et al. (2014) found that the emotion "pride" loaded differently on the factor "social value of positive emotions" across the countries in the study, implying that the value of pride was not a good indicator of the value of positive emotions in some countries. When many groups are involved, measurement noninvariances are very likely in the factor loadings or other MM parameters (Davidov et al., 2012; Zercher et al., 2015). While metric invariance must hold to ensure the comparability of structural relations, between-group differences in item intercepts and residual or "unique" variances (see the Method section) do not invalidate comparisons of the structural relations, as long as they are included in the model. If MM parameters are mistakenly restricted to be equal across groups, structural relations are incorrectly estimated (Chen, 2008; Guenole & Brown, 2014; Pokropek et al., 2019).

[1] Multilevel regression has been commonly used to compare variables' relations among many groups (e.g., Brandt et al., 2021; Rai et al., 2013), but it is beyond the scope of this article, since we focus on SEM methods.

[2] Throughout this article, the term "clusters" refers to latent classes and not to the higher-level units in multilevel modeling.

[3] Disagreement exists about how many parameters should be invariant for the sake of comparability. Some authors argue that a majority of the measurement parameters should be invariant (Reise et al., 1993), and others state that half (Steinmetz, 2013) or even a majority of the parameters can be noninvariant if correctly modeled (Pokropek et al., 2019).

Hence, to validly cluster groups based on the structural relations, all between-group differences in the MM should be accounted for in the model. Several mixture SEM methods have been proposed in the literature (e.g., Arminger & Stein, 1997; Vermunt & Magidson, 2005) to deal with differences in the SM and the MM, also when clustering groups as a whole (e.g., Kim et al., 2016). However, these methods constrain all SEM parameters to be equal within a cluster. In other words, both the MM (e.g., loadings and intercepts) and the SM (e.g., covariances or relations between latent variables) are constrained to be equal across groups within a cluster. This implies that the clustering would capture two sources of differences, that is, in the structural relations (of interest to the research question) and in the measurement parameters (not of interest to the research question). Therefore, these methods would need many clusters to properly capture both types of differences or fail to capture differences in the structural relations entirely when the MM differences are the most dominant.

To solve this important shortcoming, we present a new method called mixture multigroup SEM (MMG-SEM), which builds on the recently proposed mixture multigroup factor analysis (MMG-FA; De Roover, 2021; De Roover et al., 2022) and applies a mixture clustering of groups in a more flexible way. Specifically, the mixture multigroup approach allows combining cluster-specific parameters with parameters that remain group-specific, so that the clustering focuses on the parameters of interest. In De Roover (2021) and De Roover et al. (2022), MMG-FA aimed for MI by clustering the groups according to their MM parameters while setting aside differences in the SM (i.e., the SM can differ among groups in the same cluster). The novel MMG-SEM applies the mixture multigroup approach to cluster groups specifically on the structural relations while capturing measurement noninvariances with group-specific parameters.

We apply a two-step approach to estimate MMG-SEM since it is impractical and computationally inefficient to identify the measurement noninvariances across the groups while simultaneously clustering the groups on their structural relations. Specifically, we apply an extension of the recently proposed structural-after-measurement (SAM) approach (Rosseel & Loh, 2022). In the first step, multigroup confirmatory factor analysis (MG-CFA), a special case of MG-SEM, is used to estimate the partially group-specific MM parameters. In particular, noninvariant loadings, intercepts, and unique variances are specified as group-specific parameters, whereas invariant MM parameters are constrained to be equal across groups, keeping in mind that the loadings must be partially invariant to ensure comparability of the structural relations. In the second step, the structural relations and the mixture clustering based on these relations are estimated using the group-specific factor covariances resulting from the first step (see the Method section for details). By capturing the measurement noninvariances with group-specific parameters (in the first step) MMG-SEM ensures that the structural relations are the only parameters driving the clustering (in the second step).

The remainder of this article is as follows: MMG-SEM and its relation to MG-CFA and MMG-FA are presented in the Method section. This is followed by the Simulation Studies section used to evaluate the MMG-SEM's performance in terms of cluster and parameter recovery. Next, we present an Empirical Application section of MMG-SEM. Specifically, we re-analyzed the data from Kuppens et al. (2008) to compare the relationship between experienced emotions and life satisfaction across 47 countries. The article concludes with a Discussion section that highlights the key results, limitations, and directions for future research.

Method

Mixture MG-SEM

MMG-SEM uses the mixture multigroup approach, which combines group-specific and cluster-specific parameters to find clusters based on SEM parameters of interest. The original mixture multigroup method, MMG-FA (De Roover, 2021; De Roover et al., 2022; Leitgöb et al., 2022),[4] aimed to identify clusters according to the equivalence of specific measurement parameters (e.g., loadings). MMG-SEM is an extension of MMG-FA where the clustering of the groups is driven specifically by the structural relations between latent variables, whereas differences in measurement parameters are captured by group-specific parameters, and thus not by the clustering.

For its estimation, we build on the SAM approach from Rosseel and Loh (2022). SAM pertains to the two-step estimation of SEM models, where the MM is estimated in the first step, and the SM in the second step. The two-step estimation approach allows the researcher to evaluate the MM and potential measurement noninvariances before modeling the structural relations and, in the case of MMG-SEM, clustering the groups according to these relations. Furthermore, the step-wise approach simplifies the estimation of the clustering of the groups. One-step estimation methods would (re-)estimate both the measurement and structural parameters after each update of the clustering of the groups. When many groups are involved, this would be very costly in terms of computation time due to the large number of parameters. It is noteworthy that Rosseel and Loh (2022) proposed two variants of SAM: "local" and "global" SAM. In local SAM, estimating the SM in the second step only involves the structural parameters. In contrast, in global SAM, the second step still involves the MM parameters but they are fixed to values derived from the first step (for more details, see Rosseel & Loh, 2022). In this article, we apply the local SAM variant. Below, we elaborate on the two steps of local SAM in the context of MMG-SEM.

Step 1: MM With Measurement Noninvariances

The MM defines how latent variables are measured, that is, which items measure which factor and to what extent. In a multigroup setting, it is commonly estimated using MG-CFA. Consider individuals $n_g = 1, \ldots, N_g$ within groups $g = 1, \ldots, G$, items $j = 1, \ldots, J$, and factors $q = 1, \ldots, Q$. MG-CFA describes the J-dimensional vector of observed scores $\mathbf{x}_{n_g}$ of individual n_g belonging to group g by the linear equation:

$$\mathbf{x}_{n_g} = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\eta}_{n_g} + \boldsymbol{\epsilon}_{n_g}, \tag{1}$$

where $\boldsymbol{\tau}_g$ is a J-dimensional vector of group-specific intercepts, $\boldsymbol{\Lambda}_g$ denotes a $J \times Q$ matrix of group-specific factor loadings, $\boldsymbol{\eta}_{n_g}$ is a Q-dimensional vector of factor scores, and $\boldsymbol{\epsilon}_{n_g}$ is a J-dimensional vector of residuals. As can be seen from this equation, the loadings in $\boldsymbol{\Lambda}_g$ act as regression slopes. They represent the expected change in

[4] For information about its implementation, see the R package "mixmgfa" on https://github.com/KimDeRoover/mixmgfa.

$\mathbf{x}_{n_g}$ due to a one-unit change in one of the latent variable's scores in $\boldsymbol{\eta}_{n_g}$. MG-CFA imposes a priori constraints on $\boldsymbol{\Lambda}_g$ by specifying a pattern of zero and nonzero factor loadings. For group g, we assume that the factor scores are distributed according to a multivariate normal distribution MVN($\boldsymbol{\alpha}_g$, $\boldsymbol{\Phi}_g$), where $\boldsymbol{\alpha}_g$ and $\boldsymbol{\Phi}_g$ are the mean vector and covariance matrix of the factors, respectively. The scale of the latent variables can be set by standardizing the factors (i.e., fixing their variance to one) or by using a marker variable (i.e., fixing one loading to one per factor) for each group. Throughout this article, we use the marker variable approach. Lastly, we assume that $\boldsymbol{\epsilon}_{n_g}$ is distributed as MVN($\mathbf{0}$, $\boldsymbol{\Theta}_g$), where the diagonal of $\boldsymbol{\Theta}_g$ contains the residual or unique variances of the items in group g. When assuming the covariances between the factors $\boldsymbol{\eta}_{n_g}$ and the residuals $\boldsymbol{\epsilon}_{n_g}$ to be zero, it follows that the model-implied covariance matrix of group g is given by $\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}_g\boldsymbol{\Phi}_g\boldsymbol{\Lambda}_g' + \boldsymbol{\Theta}_g$. For each group, the MM is fitted by minimizing the difference between the model-implied covariance matrix $\boldsymbol{\Sigma}_g$ and the sample covariance matrix $\mathbf{S_g}$.

Given that MMG-SEM aims to compare latent variables' relations across groups, it is necessary to test for invariance of the MM—or MI—prior to performing an MMG-SEM analysis. First, one tests for configural invariance by evaluating whether the baseline model in Equation 1 holds across all groups. If the fit of this model is satisfactory (for an overview of fit indices in SEM, see Chen, 2007; Cheung & Rensvold, 2002), one can assume that the pattern of nonzero loadings in $\boldsymbol{\Lambda}_g$ (i.e., which items measure which factors) is the same for all groups. Then, one constrains the loadings to be equal across groups to examine metric invariance. Specifically, metric invariance holds if the model fit is not significantly worse than the fit of the baseline model when imposing $\boldsymbol{\Lambda}_g = \boldsymbol{\Lambda}$ for all groups (see guidelines by Rutkowski & Svetina, 2014). Similarly, scalar invariance holds when $\boldsymbol{\tau}_g = \boldsymbol{\tau}$ for all groups, and strict invariance holds when the diagonal of $\boldsymbol{\Theta}_g$ is equal for all groups. As was mentioned in the Introduction, at least full or partial metric invariance must hold to validly compare structural relations in MMG-SEM, where full metric invariance pertains to all loadings being invariant and partial metric invariance to some loadings being invariant.

Following this requirement, in Step 1 of MMG-SEM, the loadings were restricted to be equal whereas other MM parameters were allowed to differ across groups. Hence, the model-implied covariance matrix of the first step of MMG-SEM is given by

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}\boldsymbol{\Phi}_g\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_g. \tag{2}$$

Note that the $\boldsymbol{\Lambda}$ matrix may contain some group-specific loadings if necessary according to the MI assessment (i.e., in case of partial metric invariance), but we use the full invariance notation without subscript g for consistency. Since MI testing is performed prior to the MMG-SEM analysis, it is known which measurement parameters should be specified as invariant and which as noninvariant in the first step of MMG-SEM. Throughout this article, we specified the residual variances to be fully group-specific and removed the mean structure of the data (i.e., the group-specific means) by centering the observed item scores per group. The latter is equivalent to—but computationally more efficient than—specifying group-specific intercepts $\boldsymbol{\tau}_g$ (and group-specific factor means $\boldsymbol{\alpha}_g$ of zero). Surely, in empirical practice, if certain intercepts and residual variances are invariant, the MG-CFA specification in the first step of MMG-SEM can be adjusted accordingly.

Lastly, in MG-CFA, the estimation of the measurement parameters for all factors is usually done in one single multigroup factor analysis. However, for the first step of SAM, Rosseel and Loh (2022) introduced the so-called "measurement blocks," which indicate whether (subsets of) the latent variables should be estimated with separate (multigroup) factor analyses or whether all latent variables should be estimated at the same time with one single (multigroup) factor analysis. Rosseel and Loh (2022) recommended estimating the MM of each latent variable separately with measurement blocks that contain one latent variable each. If MM misspecifications (e.g., omitted residual covariance between items measuring the same factor) are not included for one of the latent variables, the separate measurement blocks will contain the effect of the misspecification (e.g., biased estimations) and prevent it from spreading to other blocks. Thus, the model becomes more robust against MM misspecifications when using separate measurement blocks. However, specifying one measurement block per latent variable is not always possible. For instance, in the case that, due to theoretical knowledge, a researcher wants to include cross-loadings or residual covariances between indicators of two factors, it would be necessary to include both factors in the same measurement block. If more than one measurement block is specified, the factor covariance matrix $\boldsymbol{\Phi}_g$ in Equation 2 is computed using the so-called "mapping matrix" $\mathbf{M}$ from Rosseel and Loh (2022), which would differ across groups in the case of MMG-SEM.

The group-specific covariances matrices $\boldsymbol{\Phi}_g$ resulting from Step 1 serve as the input for the second step of MMG-SEM. To avoid confusion, from now on, $\boldsymbol{\Phi}_g$ will be used with the superscript $s1$ (i.e., $\boldsymbol{\Phi}_g^{s1}$) to identify it as a group-specific factor covariance matrix of Step 1. For the estimation of the MM parameters via maximum likelihood (ML), we use the R-package "lavaan" (Rosseel, 2012). For more details on the implementation of Step 1, the reader is referred to Appendix A.

Step 2: SM With Mixture Clustering of the Groups

In the second step, MMG-SEM estimates the SM and the mixture clustering of the groups based on the structural relations. The SM is conditional on the membership of group g to cluster k, denoted as $\mathbf{z}_{gk}$. This cluster membership takes on a value of either 1 or 0, indicating whether group g belongs to cluster k or not, respectively. The formal definition of the SM is as follows:

$$[\boldsymbol{\eta}_{n_g}|\mathbf{z}_{gk} = 1] = \boldsymbol{\nu}_g + \mathbf{B}_k\boldsymbol{\eta}_{n_g} + \boldsymbol{\zeta}_{n_{gk}}, \tag{3}$$

where $\boldsymbol{\nu}_g$ is a vector of group-specific factor intercepts, $\mathbf{B}_k$ is a nonsymmetric $Q \times Q$ matrix containing the unstandardized cluster-specific regression coefficients between latent variables where the diagonal elements are zero, and $\boldsymbol{\zeta}_{n_g}$ is a Q-dimensional vector of disturbances (i.e., residuals) of the regressions among latent variables.[5] It should be noted that while the true cluster membership $\mathbf{z}_{gk}$ is unknown (latent), an estimation $\hat{\mathbf{z}}_{gk}$ can be obtained, which is a membership probability ranging from 0 to 1 reflecting the classification certainty. For more details on the computation of $\hat{\mathbf{z}}_{gk}$, see

[5] The $\boldsymbol{\zeta}_{n_{gk}}$ pertain to the residuals due to a regression of a factor on another factor. They differ from the residual variances in $\boldsymbol{\Theta}_g$ of the MM, which result from the regression of an observed item on a factor.

Appendix A. If we assume that $\mathrm{E}(\boldsymbol{\zeta}_{n_g}) = 0$ and $\mathbf{Cov}(\boldsymbol{\zeta}_{n_g}) = \boldsymbol{\Psi}_g$, then the model-implied factor covariance matrix in Step 2 is given by

$$\boldsymbol{\Phi}_{gk} = (\mathbf{I} - \mathbf{B}_k)^{-1}\boldsymbol{\Psi}_{gk}(\mathbf{I} - \mathbf{B}_k)^{-1'}. \tag{4}$$

The SM is then fitted by minimizing the difference between the group-specific factor covariance matrices $\boldsymbol{\Phi}_g^{s1}$ obtained in Step 1 (Equation 2), and their model-implied reconstruction $\boldsymbol{\Phi}_{gk}$ (Equation 4). To obtain the clustering of the groups based on the regression parameters in $\mathbf{B}_k$, MMG-SEM assumes that the vector of factor scores η_{n_g} is sampled from a mixture of K multivariate normal distributions and that all factor scores of group g—gathered in a matrix $\mathbf{H}_g$ of factor scores—are sampled from the same distribution. Specifically, the formal definition for group g is the following:

$$\begin{aligned} f(\mathbf{H}_g;\vartheta) &= \sum_{k=1}^{K} \pi_k f_{gk}(\mathbf{H}_g;\ \vartheta_{gk}) \\ &= \sum_{k=1}^{K} \pi_k \prod_{n_g=1}^{N_g} \mathrm{MVN}(\boldsymbol{\eta}_{n_g};\boldsymbol{\alpha}_g, \boldsymbol{\Phi}_{gk}), \end{aligned} \tag{5}$$

for which f is the total population density function, ϑ is the set of population parameters, π_k is the prior probability of a group g belonging to cluster k (where $\sum_{k=1}^{K} \pi_k = 1$), f_{gk} is the density function of the group g in the kth cluster, and ϑ_{gk} is its corresponding set of parameters. Specifically, f_{gk} is a multivariate normal distribution where $\boldsymbol{\Phi}_{gk}$ and α_g are the factors' covariance matrix and mean vector, respectively. The covariance matrix is decomposed as indicated in Equation 4, and the factor means α_g are equal to 0 throughout this article.

If we insert Equation 4 into Equation 2—that is, if we combine the MM and SM—we obtain the following model-implied covariance matrix of the observed data for group g in cluster k:

$$\boldsymbol{\Sigma}_{gk} = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B}_k)^{-1}\boldsymbol{\Psi}_{gk}(\mathbf{I} - \mathbf{B}_k)^{-1'}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_g. \tag{6}$$

Please note that the model-implied covariance matrix $\boldsymbol{\Sigma}_{gk}$ is composed of (partially) invariant factor loadings $\boldsymbol{\Lambda}$, cluster-specific regression coefficients $\mathbf{B}_k$, group-and-cluster-specific residual factor covariances $\boldsymbol{\Psi}_{gk}$, and group-specific residual variances $\boldsymbol{\Theta}_g$. The structural relations are the only cluster-specific parameters, so the clustering is driven solely by them.

To understand the group-and-cluster-specific components of the residual factor covariances in $\boldsymbol{\Psi}_{gk}$ consider its partitioned matrix definition below:

$$\boldsymbol{\Psi}_{gk} = \begin{bmatrix} \boldsymbol{\Psi}_g^{\mathrm{exog}} & 0 & 0 \\ 0 & \boldsymbol{\Psi}_{gk}^{\mathrm{endog1}} & 0 \\ 0 & 0 & \boldsymbol{\Psi}_{gk}^{\mathrm{endog2}} \end{bmatrix}, \tag{7}$$

where $\boldsymbol{\Psi}_g^{\mathrm{exog}}$ is the group-specific nondiagonal covariance matrix of the exogenous factors, $\boldsymbol{\Psi}_{gk}^{\mathrm{endog1}}$ is the group-and-cluster-specific diagonal residual variance matrix of the endogenous factors that act both as dependent and independent variables in the SM (i.e., mediators; referred to as "endogenous-1"), and $\boldsymbol{\Psi}_{gk}^{\mathrm{endog2}}$ denotes the group-and-cluster-specific nondiagonal residual covariance matrix of the endogenous factors that are purely dependent variables ("endogenous-2").[6] The covariance matrix of exogenous variables $\boldsymbol{\Psi}_g^{\mathrm{exog}}$ was specified as group-specific as it should not contribute to the (mis)fit of the model (i.e., it should not affect the clustering). In contrast, the residual covariances of the endogenous variables was defined as group-and-cluster-specific due to its dependence on the cluster-specific regression coefficients $\mathbf{B}_k$. The default diagonal (i.e., no residual covariances) and nondiagonal (i.e., allowing residual covariances) nature of $\boldsymbol{\Psi}_{gk}^{\mathrm{endog1}}$ and $\boldsymbol{\Psi}_{gk}^{\mathrm{endog2}}$, respectively, was defined considering common assumptions in the SEM framework. That is, we assume that the relations between the mediator variables in $\boldsymbol{\Psi}_{gk}^{\mathrm{endog1}}$ are fully explained by the model. In contrast, the relations between purely endogenous factors in $\boldsymbol{\Psi}_{gk}^{\mathrm{endog2}}$ are not necessarily fully explained, which is why we allow them to be correlated (similar to seemingly unrelated regression; Zellner, 1962). Nevertheless, as assumptions can differ from case to case, MMG-SEM allows removing the residual covariances between endogenous-2 factors if the researcher requires it. Note that the group-and-cluster-specific nature of the residual factor covariances is the default for MMG-SEM. However, if the researcher requires it (e.g., if these residual (co)variances are considered to be part of the relations one wants to cluster on), $\boldsymbol{\Psi}_{gk}^{\mathrm{endog1}}$ and $\boldsymbol{\Psi}_{gk}^{\mathrm{endog2}}$ can be specified as purely cluster-specific.

Note that between-group differences in the exogenous factor variances $\boldsymbol{\Psi}_g^{\mathrm{exog}}$ may exist, even among groups in the same cluster. This is not an issue for the scale of the cluster-specific regression coefficients of the exogenous factors, since, in each group, the factors' scale is set by restricting the loading of the same marker variable to one, so that, in each cluster, these regression parameters pertain to the effect of a one-unit change of an exogenous factor on an endogenous factor. A one-unit change then has the same meaning in all groups, so that differences in exogenous variances are merely differences in variability concerning these factors.

Model Estimation

The unknown parameters ϑ of Step 2 in Equation 5 are estimated by means of ML estimation. Specifically, the estimation of the SM, including the mixture model, maximizes the following log-likelihood function:

$$\begin{aligned} \mathrm{logL}_{\boldsymbol{\eta}} &= \log\left(\prod_{g=1}^{G}\sum_{k=1}^{K}\pi_k\left(\frac{1}{(2\pi)^{Q/2}|\boldsymbol{\Phi}_{gk}|^{1/2}}\exp\left(-\frac{1}{2}tr\left(\boldsymbol{\Phi}_g^{s1}\boldsymbol{\Phi}_{gk}^{-1}\right)\right)\right)^{N_g}\right) \\ &= \sum_{g=1}^{G}\log\left(\sum_{k=1}^{K}\pi_k\left(\frac{1}{(2\pi)^{Q/2}|\boldsymbol{\Phi}_{gk}|^{1/2}}\exp\left(-\frac{1}{2}tr\left(\boldsymbol{\Phi}_g^{s1}\boldsymbol{\Phi}_{gk}^{-1}\right)\right)\right)^{N_g}\right), \end{aligned} \tag{8}$$

where $\boldsymbol{\Phi}_g^{s1}$ is the group-specific factor covariance matrix from Step 1 (Equation 2), and $\boldsymbol{\Phi}_{gk}$ is the group-and-cluster-specific factor covariance matrix from Step 2 (Equation 4). As is common in mixture modeling, we applied an expectation-maximization (EM; Dempster et al., 1977) algorithm to optimize this log-likelihood function. To avoid local maxima, a multistart procedure was implemented in Step 2, where each start uses a random hard partition (i.e., each group belongs fully to one cluster, with a probability of one). The best solution resulting from the multiple starts (i.e., with the highest $\mathrm{logL}_{\boldsymbol{\eta}}$) is selected as the final result. The MMG-SEM algorithm is detailed in Appendix A and has been implemented in R (https://github.com/AndresFPA/MMG-SEM).

[6] The gk subscript of $\boldsymbol{\Psi}_{gk}^{\mathrm{endog1}}$ and $\boldsymbol{\Psi}_{gk}^{\mathrm{endog2}}$ implies that the residual (co)variance matrices are estimated for each group-cluster combination.

The log-likelihood function maximized in Equation 8 pertains to the structural parameters from Step 2. If we consider both the MM and the SM—that is, combining Steps 1 and 2—, we obtain the following log-likelihood function of the full model:

Equation 9 (see below)

where $\boldsymbol{\mu}_g$ and $\boldsymbol{\Sigma}_{gk}$ are the mean vector and model-implied covariance matrix of the observed items, respectively. The covariance matrix $\boldsymbol{\Sigma}_{gk}$ is decomposed as in Equation 6.

Model Selection

In empirical research, the number of clusters underlying the data is unknown, and the appropriate number of clusters K needs to be selected. Often, the Bayesian information criterion (BIC; Schwarz, 1978) is recommended to solve this problem (e.g., Nylund et al., 2007). The BIC is a model selection criterion that combines the likelihood of the model with a penalty based on its number of parameters. It is defined as:

$$\text{BIC} = -2\text{logL} + P\log(\text{SS}), \tag{10}$$

where logL is the log-likelihood, SS is the sample size, and P is the number of free parameters in the model. P is the sum of the number of mixing proportions (minus one restriction), the number of cluster-specific regressions coefficients, the number of group-specific exogenous factor covariances, the number of group-and-cluster-specific "endogenous-1" factor variances, the number of group-and-cluster-specific "endogenous-2" factor covariances,[7] the number of (partially) invariant loadings (minus Q fixed loadings due to factor scaling, potentially accounting for some loadings being group-specific), and the number of group-specific unique variances.

Usually, the total number of observations N is used as the SS in Equation 10. Alternatively, it has been suggested to replace N with the number of groups G if the model selection issue pertains to group-level clusters (Lukočienė & Vermunt, 2009). The model selection performance of the BIC using N (BIC_N) and using G (BIC_G) was already tested in the context of the mixture multigroup approach (De Roover, 2021; De Roover et al., 2022), indicating a superior performance for the BIC_G. We assume this to hold for MMG-SEM as well.

Finally, it is worth noting that the MMG-SEM model is fitted by minimizing the loglikelihood of the structural parameters $\text{logL}_{\boldsymbol{\eta}}$ (Equation 8), whereas the observed data loglikelihood logL (Equation 9) is used for the model selection. To reconstruct logL, we use the measurement parameters from Step 1 and the final structural parameters from Step 2. Using the complete data loglikelihood, we ensure that the model selection is based on the complete model and how that fits the observed data at hand.

Simulation Studies

Two simulation studies are presented in this section. Both simulations evaluated the performance of MMG-SEM in terms of cluster recovery and parameter recovery when the number of clusters is known. In Simulation Study 1, we assessed the performance of MMG-SEM when measurement noninvariances were captured by group-specific parameters. In Simulation Study 2, we investigated how ignoring measurement noninvariances affects the recovery of clusters and parameters. The code behind the simulations is publicly available on GitHub and can be accessed at https://github.com/AndresFPA/MMG-SEM_Simulation.

Simulation Study 1

Design

The goal of Simulation Study 1 was to assess the cluster and parameter recovery by means of MMG-SEM when the true number of clusters is known and when the measurement noninvariances are known and captured by group-specific parameters. Specifically, we evaluated MMG-SEM through a Monte Carlo simulation wherein we manipulated eight factors that likely influence the performance of the method. The factors and their corresponding levels are below:

1. Size of regression parameters β (three levels): .2, .3, and .4.
2. Number of groups G (three levels): 12, 24, and 48.
3. Within-group sample size N_g (three levels): 50, 100, and 200.
4. Number of clusters K (two levels): 2 and 4.
5. Cluster size (two levels): balanced and unbalanced.
6. Reliability (two levels): low (0.4) and high (0.6).
7. Proportion of noninvariant groups (three levels): 0.25, 0.5, and 0.75.
8. Loading noninvariance size (two levels): 0.4 and 0.6.

Note that factors 7 and 8 refer specifically to measurement noninvariance added to the loadings of the model. This resulted in a total of 3 (size of regression parameters) × 3 (number of groups) × 3 (within-group sample size) × 2 (number of clusters) × 2 (cluster size) × 2 (reliability) × 3 (proportion of noninvariant groups) × 2 (noninvariance size) = 1,296 data generation conditions. For all 1,296 conditions, 50 replications were simulated according to the data generation procedure described below, using R Version 4.1.1 (R Core Team, 2022). In total, 64,800 data sets were generated. All data sets were analyzed using the MMG-SEM algorithm described in Appendix A with 20 random starts.

[7] We only count one set of endogenous covariances for each group, given we assume each group belongs to only one cluster. The endogenous covariances (from the clusters the groups do not belong to) are nuisance parameters.

$$\begin{aligned}\text{logL} &= \log\left(\prod_{g=1}^{G}\sum_{k=1}^{K}\pi_k\prod_{n_g=1}^{N_g}\frac{1}{(2\pi)^{J/2}|\boldsymbol{\Sigma}_{gk}|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}_{n_g}-\boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_{gk}^{-1}(\mathbf{x}_{n_g}-\boldsymbol{\mu}_g)\right)\right)\\ &= \sum_{g=1}^{G}\log\left(\sum_{k=1}^{K}\pi_k\prod_{n_g=1}^{N_g}\frac{1}{(2\pi)^{J/2}|\boldsymbol{\Sigma}_{gk}|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}_{n_g}-\boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_{gk}^{-1}(\mathbf{x}_{n_g}-\boldsymbol{\mu}_g)\right)\right),\end{aligned} \tag{9}$$

Data Generation

The data for the simulation were generated according to a SEM model with four latent variables. The model is depicted in Figure 1. The latent variables *F*1 and *F*2 served as exogenous variables, while the other two were endogenous variables. The endogenous variable *F*3 was modeled as an independent and dependent variable at the same time ("Endogenous-1"), and the remaining latent variable *F*4 was a dependent only variable ("Endogenous-2"). All latent variables were measured by five items, for a total of 20 observed variables.

The number of groups G ranged from 12 to 48 depending on the "number of groups" factor, whereas the group sizes N_g were 50, 100, or 200 depending on the manipulation of the "within-group sample size" factor. Similarly, the number of clusters K was two or four depending on the factor "number of clusters," and the number of groups per cluster was determined by the "cluster size" factor. Specifically, for the balanced condition, all clusters contained the same number of groups. For example, for $G = 24$ and $K = 4$, all clusters contained six groups. In contrast, for the unbalanced condition, we generated the first cluster as a large cluster containing 75% of the groups, while the remaining clusters were generated as smaller equally sized clusters for the remaining 25% of the groups. For $G = 24$ and $K = 4$, the first cluster would then gather 18 groups, and each of the remaining three clusters would contain only two groups.

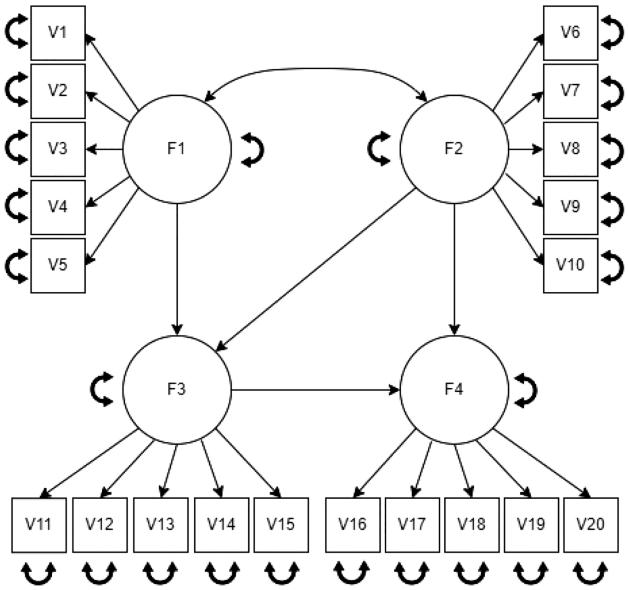
The means $\boldsymbol{\mu}_g$ were zero for all groups, and the sample covariance matrix $\boldsymbol{\Sigma}_{gk}$ was generated following Equation 6. Thus, the parameters in Equation 6 had to be defined. The regression parameters in $\mathbf{B}_k$ were generated as depicted in Figure 2. The size of β_1, β_2, β_3, and β_4 depended on the manipulated factor "size of regression parameters," and for each cluster, one of them was set to zero. Since the regression parameter set to zero was the main difference between the clusters, the size of the regression parameters also determines the size of the difference between the clusters. For each pair of clusters, the difference between them pertains to two regression coefficients. In the case of two clusters, the third and fourth models in Figure 2 are not applicable. The group-specific variances for the exogenous latent variables *F*1 and *F*2 were sampled from the uniform distribution $U(0.75, 1.25)$, and their covariance was sampled from $U(-0.3, 0.3)$ for all groups. The total variances of the endogenous factors *F*3 and *F*4 were sampled from $U(0.75, 1.25)$. Their residual variances depended on the regression coefficients. For instance, when the total variance of *F*3 in group g was σ_g, *F*3's residual variance for group g in cluster k was $\sigma_g - \left(\beta_2^2 \mathrm{Var}(F1) + \beta_3^2 \mathrm{Var}(F2) + 2\beta_2\beta_3 \mathrm{Cov}(F1, F2)\right)$.

The loadings matrix $\boldsymbol{\Lambda}$ was initially set as invariant across all groups. To obtain total variances per item that vary around 1, the nonzero values of $\boldsymbol{\Lambda}$ were set to $\sqrt{.4}$ with unique variances $\boldsymbol{\Theta}_g$ sampled from $U(0.50, 0.70)$ when the reliability level was low. In contrast, when the reliability level was high, the loadings in $\boldsymbol{\Lambda}$ were $\sqrt{.6}$ and the diagonal of $\boldsymbol{\Theta}_g$ was drawn from $U(0.30, 0.50)$. The unique variances were sampled per group, which allowed them to differ across groups.

To evaluate whether and how loading noninvariances affect the performance of MMG-SEM, we then added between-group differences to the loadings. First, we selected the groups affected by the noninvariance according to the manipulated factor "proportion of noninvariant groups." For instance, if the proportion of noninvariant groups was 0.25 and $G = 12$, we added the non-invariance to $12 \times 0.25 = 4$ randomly selected groups, whereas the loadings were unaltered for the remaining eight groups. Second, the differences were added to the second and third nonzero loadings of each latent variable (the first loading is fixed to 1 to set the scale of the factors). The size of the noninvariance was defined based on the "noninvariance size" manipulation. Specifically, loading differences take on a value of 0.4 or 0.6. To maintain the average loading value for each latent variable, we added the noninvariance to one loading and subtracted it from the other modified loading. An example of an invariant and noninvariant $\boldsymbol{\Lambda}$ matrix can be seen in Table 1.

Finally, after having defined all the necessary parameters, the data were sampled from the multivariate normal distribution $\mathrm{MVN}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_{gk})$ per group.

Figure 1
The Model Used for the Data Generation



Note. *F*1 and *F*2 are the exogenous variables, *F*3 is the dependent and independent at the same time ("Endogenous-1"), and *F*4 is a dependent only variable ("Endogenous-2"). V1–V20 = observed variables 1–20.

Results

Computation Times. On a workstation consisting of a CPU Intel Xeon Gold 6248R with a clock frequency of 3.00 GHz and RAM of 512 GB, the average computation time of MMG-SEM with 20 random starts was 51.4 s. Note that the computation time greatly depended on the conditions of the simulation. The overall minimum was 6 s in an "easy" condition with fewer groups (e.g., $K = 2$, $G = 12$, $\beta = .4$), and the maximum was 30 min in a "hard" condition with more groups (e.g., $K = 4$, $G = 48$, $\beta = .2$).

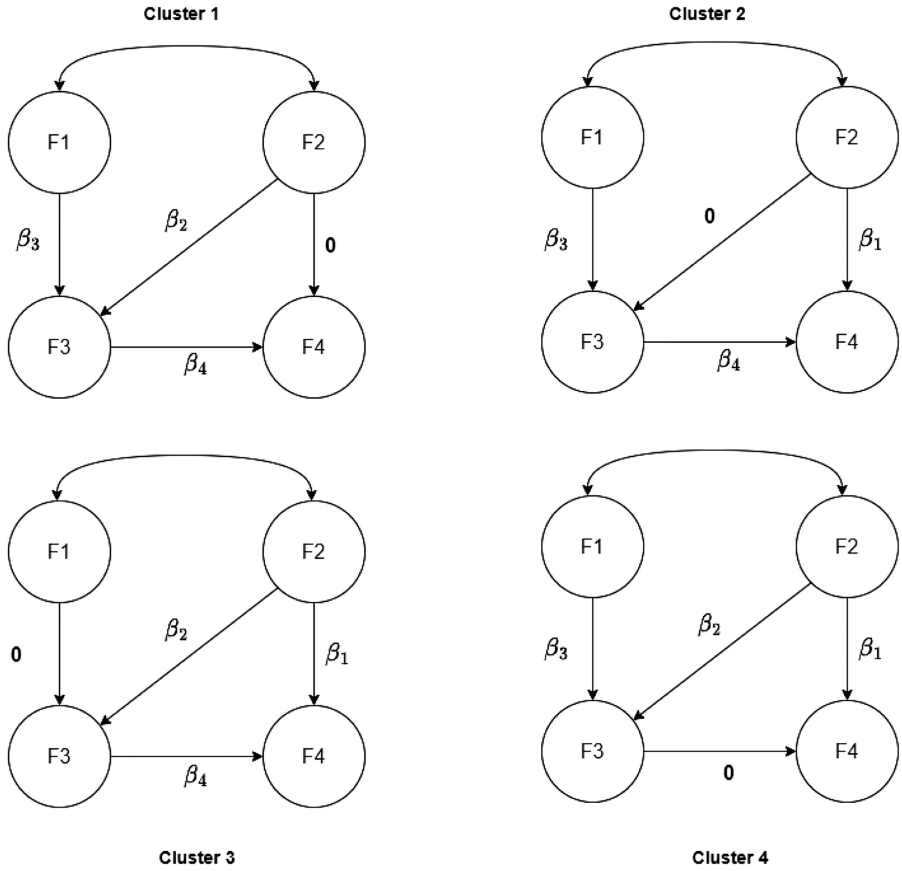
Recovery of the MM. We evaluated the recovery of the MM parameters in Step 1 of MMG-SEM using the root mean squared error (RMSE). For the factor loadings, we computed RMSE as follows:

$$\mathrm{RMSE}_\Lambda = \sqrt{\frac{\sum_{g=1}^{G}\sum_{j=1}^{J}(\lambda_{gj} - \hat{\lambda}_{gj})^2}{GJ}}, \tag{11}$$

where λ_{gj} is the true factor loading of item j on the factor it is measuring and $\hat{\lambda}_{gj}$ is its corresponding estimation. We excluded the first loading of

Figure 2

Zero and Nonzero Regression Parameters Between the Latent Variables Depending on the Cluster



Note. *F*1 and *F*2 are the exogenous variables, *F*3 is the dependent and independent at the same time ("Endogenous-1"), and *F*4 is a dependent only variable ("Endogenous-2").

each factor from the computation of RMSE_Λ since they are fixed to one. For the item residual variances, RMSE was:

$$\text{RMSE}_\Theta = \sqrt{\frac{\sum_{g=1}^{G}\sum_{j=1}^{J}(\theta_{gj} - \hat{\theta}_{gj})^2}{GJ}}, \tag{12}$$

where θ_{gj} is the true residual variance for item j in the diagonal of Θ_g, and $\hat{\theta}_{gj}$ is its corresponding estimation. On average, the results indicated an excellent recovery for both the factor loadings (i.e., $\text{RMSE}_\Lambda < .001$, $SD < 0.001$) and the item residual variances (i.e., $\text{RMSE}_\Theta = .006$, $SD = 0.003$), and were hardly affected by the manipulated conditions. All following reported findings in this section pertain to Step 2 of MMG-SEM, focusing on the parameters associated with the SM and its clustering.

Sensitivity to Local Maxima. To evaluate how often our algorithm converged at a local maximum, we compared the final logL_η (i.e., the best random start from Step 2; see Appendix A) obtained for each of the 64,800 data sets against the global ML solution. Since the logL_η from the global maximum is unknown due to sampling fluctuations, we obtained an approximation or "proxy" to this solution by starting the algorithm from the true partition. The solution resulting from 20 random starts was thus a local maximum when its logL_η was smaller than the one from the proxy. To exclude differences due to calculation precision, we considered a solution a local maximum when its logL_η was more than 0.001 lower than that of the proxy.

The results showed that 12.46% (8,076) of all 64,800 data sets converged to a local maximum. The simulation factors with the largest effect on the sensitivity to local maxima were the cluster sizes and the number of clusters. Regarding the cluster size factor, all 8,076 local maxima occurred in unbalanced cluster sizes. Similarly, 7,417 out of the 8,076 local maxima occurred when the number of clusters was four. Other factors that affected the sensitivity to local maxima were the within-group sample size and the size of the regression parameters. The percentage of local maxima data sets were 17.1%, 13.1%, and 7.1% when the N_g was 50, 100, and 200, respectively. Showing a similar trend, the percentage of local maxima data sets were 16.6%, 13.8%, and 6.9% when the β was 0.2, 0.3, and 0.4, respectively.

To gain a deeper understanding of the sensitivity, we also evaluated, for each data set, how many of the 20 random starts converged to a local maximum. On average, 24.7% of the random starts converged to a local maximum. Not surprisingly, the number of clusters and the cluster size were again the most influential factors. Specifically, on average, only 2.6% of the random starts converged to a local maximum when $K = 2$, whereas this number increased to 46.7% when

Table 1
Invariant and Noninvariant Λ Matrices Example

	Invariant Λ				Noninvariant Λ			
Variant	F1	F2	F3	F4	F1	F2	F3	F4
V1	1	0	0	0	1	0	0	0
V2	$\sqrt{.6}$	0	0	0	$\sqrt{\mathbf{1}}$	0	0	0
V3	$\sqrt{.6}$	0	0	0	$\sqrt{\mathbf{.2}}$	0	0	0
V4	$\sqrt{.6}$	0	0	0	$\sqrt{.6}$	0	0	0
V5	$\sqrt{.6}$	0	0	0	$\sqrt{.6}$	0	0	0
V6	0	1	0	0	0	1	0	0
V7	0	$\sqrt{.6}$	0	0	0	$\sqrt{\mathbf{1}}$	0	0
V8	0	$\sqrt{.6}$	0	0	0	$\sqrt{\mathbf{.2}}$	0	0
V9	0	$\sqrt{.6}$	0	0	0	$\sqrt{.6}$	0	0
V10	0	$\sqrt{.6}$	0	0	0	$\sqrt{.6}$	0	0
V11	0	0	1	0	0	0	1	0
V12	0	0	$\sqrt{.6}$	0	0	0	$\sqrt{\mathbf{1}}$	0
V13	0	0	$\sqrt{.6}$	0	0	0	$\sqrt{\mathbf{.2}}$	0
V14	0	0	$\sqrt{.6}$	0	0	0	$\sqrt{.6}$	0
V15	0	0	$\sqrt{.6}$	0	0	0	$\sqrt{.6}$	0
V16	0	0	0	1	0	0	0	1
V17	0	0	0	$\sqrt{.6}$	0	0	0	$\sqrt{\mathbf{1}}$
V18	0	0	0	$\sqrt{.6}$	0	0	0	$\sqrt{\mathbf{.2}}$
V19	0	0	0	$\sqrt{.6}$	0	0	0	$\sqrt{.6}$
V20	0	0	0	$\sqrt{.6}$	0	0	0	$\sqrt{.6}$

Note. The size of the noninvariance is 0.4. The first loading of each factor is fixed to 1 to set the scale of the factors. V1–V20 = observed variables 1–20; F1–F4 = Factors 1–4.

Table 2
Average Cluster Recovery Results per Level of Each Manipulated Factor

Factor	Level	ARI	%CC	fARI
β size	.2	0.718 (0.43)	0.614 (0.49)	0.653 (0.42)
	.3	0.967 (0.10)	0.859 (0.35)	0.957 (0.11)
	.4	0.984 (0.06)	0.931 (0.25)	0.983 (0.07)
G	12	0.894 (0.28)	0.832 (0.37)	0.873 (0.29)
	24	0.890 (0.28)	0.802 (0.40)	0.863 (0.29)
	48	0.886 (0.28)	0.769 (0.42)	0.856 (0.30)
N_g	50	0.713 (0.42)	0.609 (0.49)	0.658 (0.41)
	100	0.960 (0.12)	0.865 (0.34)	0.944 (0.15)
	200	0.997 (0.02)	0.929 (0.26)	0.991 (0.03)
K	2	0.947 (0.22)	0.946 (0.23)	0.915 (0.25)
	4	0.833 (0.32)	0.656 (0.47)	0.814 (0.32)
Cluster size	Balanced	0.944 (0.23)	0.943 (0.23)	0.909 (0.26)
	Unbalanced	0.836 (0.32)	0.659 (0.47)	0.820 (0.32)
Reliability	High	0.890 (0.28)	0.801 (0.40)	0.864 (0.29)
	Low	0.890 (0.28)	0.801 (0.40)	0.864 (0.29)
Noninv groups	0.25	0.890 (0.28)	0.801 (0.40)	0.864 (0.29)
	0.50	0.890 (0.28)	0.801 (0.40)	0.864 (0.29)
	0.75	0.890 (0.28)	0.801 (0.40)	0.865 (0.29)
Noninv size	0.4	0.890 (0.28)	0.801 (0.39)	0.865 (0.29)
	0.6	0.890 (0.28)	0.801 (0.39)	0.864 (0.29)
Total	—	0.890 (0.28)	0.801 (0.40)	0.864 (0.29)

Note. Results are shown for the ARI, the proportion of data sets with perfectly cluster recovery (%CC), and the fARI. The standard deviation is shown between brackets. β size is the size of (difference in) regression parameters, *G* is the number of groups, N_g is the within-group sample size, *K* is the number of clusters, Noninv *G* is the proportion of noninvariant groups. %CC is the proportion of perfect cluster recovery. CC = correct clustering; ARI = adjusted rand index; fARI = Frobenious adjusted rand index.

$K = 4$. Similarly, when the cluster sizes were balanced, only 6% of the random starts ended up in local maxima, compared to 43.3% when they were unbalanced. These results follow a similar trend to that of previous research on local optima in clustering (De Roover et al., 2022; Shireman et al., 2016; Steinley, 2006).

To conclude, MMG-SEM's sensitivity to local maxima was influenced the most by within-cluster sample size and cluster separation. The cluster separation was determined by the size of the regression parameters, whereas the within-cluster sample size depended on the number of clusters, the cluster size and the within-group sample size.

Cluster Recovery. For cluster recovery, the adjusted rand index (ARI; Hubert & Arabie, 1985), and Frobenious ARI (fARI; Andrews et al., 2022) were used. The ARI aims to compare two partitions, and takes on a value of 1 for complete agreement, and 0 when the agreement between the partitions is as expected by chance (i.e., the expected agreement between two random partitions). Thus, the ARI can take on negative values. Note that, for computing the ARI, we transformed the probabilistic cluster memberships into a hard partition by assigning each group to the cluster for which its classification probability was the highest.

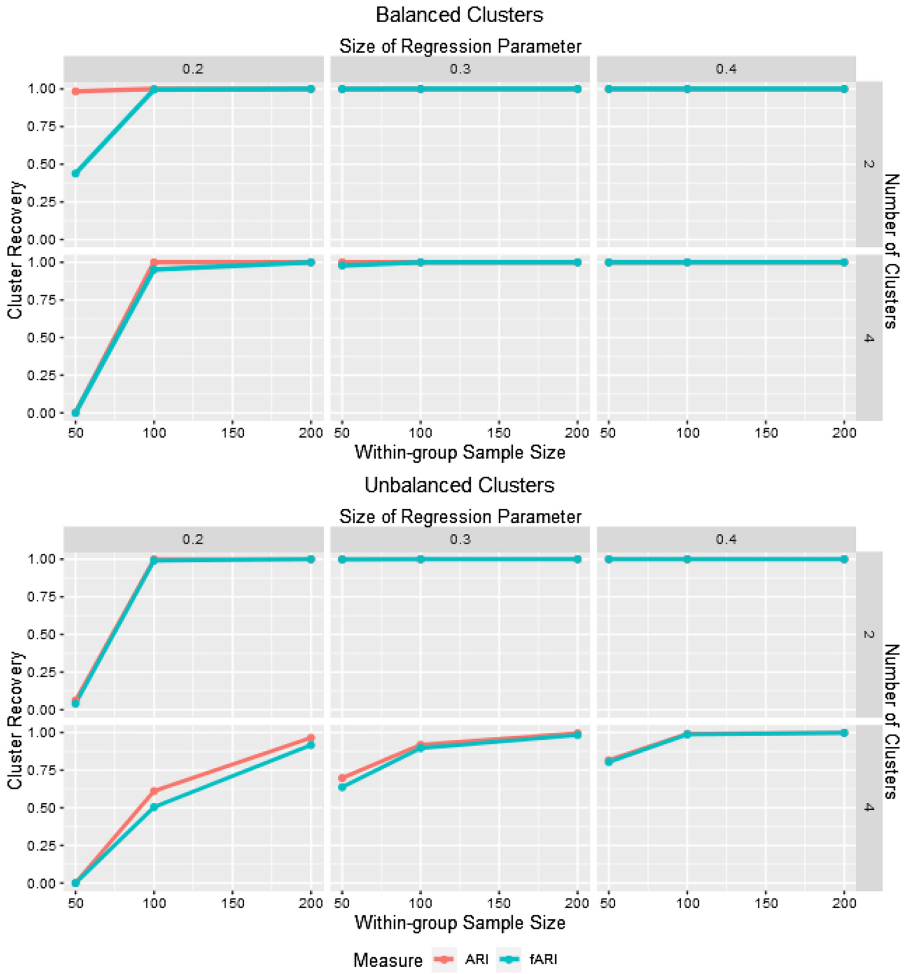
Table 2 shows the main effects of the manipulated factors on the ARI. On average, across all simulated conditions, the ARI took on a value of 0.89. Furthermore, the ARI was equal to 1 (i.e., perfect cluster recovery) for 80.1% of all generated data sets (see %CC in Table 2). Thus, overall, MMG-SEM performed well in terms of recovering the underlying clustering of the groups. The "within-group sample size" and "size of regression parameters" manipulations showed the largest effect on the cluster recovery. The larger the sample size, the better the performance of MMG-SEM. For $N_g = 50$, the ARI was 0.713 on average and an ARI of 1 was found for 60.8% of the data sets. In contrast, when $N_g = 200$, the average ARI was 0.997, close to perfect cluster recovery, and the ARI was 1 for 92.9% of all data sets. Similarly, in case of a larger difference between the clusters, it was easier for MMG-SEM to recover the clustering. When β was 0.2, 0.3, and 0.4, the average ARI, was 0.719, 0.968, and 0.984, respectively. Note that a minor increase from 0.2 to 0.3 dramatically improved the performance of MMG-SEM. The "number of clusters" and "cluster size" manipulation affected the cluster recovery to a lesser extent. Specifically, a larger number of clusters worsened the cluster recovery, as did unbalanced cluster sizes. These results are in line with previous research on clustering performance (De Roover, 2021; De Roover et al., 2022; Steinley, 2006; Steinley & Brusco, 2011), which found that the cluster recovery (i.e., ARI) depended on the sample size, number of clusters, and cluster size.

The interaction between these manipulated factors can be seen in Figure 3. The plots show a relevant interaction between the within-group sample size and the size of the regression coefficients. Clearly, MMG-SEM showed its worst performance when $N_g = 50$ and $\beta = .2$. In contrast, increasing N_g to 100 or β to 0.3, already led to a dramatic improvement in the performance (i.e., an ARI of at least 0.75). When both $N_g = 100$ and $\beta = .3$, the ARI was above 0.90.

The fARI, an extension of the ARI, allows comparing "fuzzy" partitions where each group is not solely assigned to one cluster. Specifically, in MMG-SEM (and other mixture models), each group has a probability of belonging to each cluster. Unlike the ARI, fARI provides insight into whether misclassifications of groups co-occur with classification uncertainty (i.e., whether misclassification occurs with a probability far below or close to one). The fARI results were similar, but slightly lower than the ARI results. As can be seen in Table 2, the average value of fARI was 0.864, and it was affected by the manipulated factors in the same

Figure 3

Cluster Recovery in Function of Within-Group Sample Size, Number of Clusters, Size of Regression Parameters, and Cluster Size



Note. Top: Balanced cluster size conditions. Bottom: Unbalanced cluster size conditions. ARI = adjusted rand index; fARI = Frobenious adjusted rand index. See the online article for the color version of this figure.

way as the ARI. Figure 3 presents the fARI results for the interaction between the most influential manipulated factors. Similar to the ARI, the fARI presented its lowest values when $N_g = 50$ and $\beta = .2$, and improved rapidly when $N_g = 100$ or $\beta = .3$ (i.e., to at least 0.7). When both $N_g = 100$ and $\beta = .3$, the fARI was above 0.85 on average. However, in the case of $N_g = 50$, $\beta = .2$, $K = 2$, and balanced cluster sizes, the fARI results were notably worse than the ARI results, as shown in Figure 3. After further inspection, we noticed that in such difficult conditions, MMG-SEM recovers the clusters with a high classification uncertainty (i.e., estimated cluster memberships $\hat{\mathbf{z}}_{gks}$ around 0.7), which was reflected in the fARI results.

Regression Parameter Recovery. For the recovery of the regression parameters, we decided to use the RMSE. It is defined as follows:

$$\text{RMSE}_{\beta} = \sqrt{\frac{\sum_{k=1}^{K} (\beta_k - \hat{\beta}_k)^2}{K}}, \tag{13}$$

where K is the number of clusters, β_k is the true regression parameter in cluster k, and $\hat{\beta}_k$ is the estimated regression coefficient in that same cluster. Note that "label switching" (i.e., the labels of the true and estimated clusters can differ) is a common issue in mixture models as the cluster labels are arbitrary (Tueller et al., 2011). To account for this, the obtained clusters were matched with the true clusters by maximizing the RMSE_{β}. Previous research has shown that the bias caused by the measurement noninvariances on the regression estimations depends on the type of latent variable involved (Guenole & Brown, 2014). In other words, the bias on a regression depends on whether it affects a mediator or a purely endogenous variable. Thus, the RMSE values are reported for each regression coefficient.

The main effects of the manipulated factors on the parameter-specific RMSE values are shown in Table 3. On average, the results were good for all four regression parameters (i.e., $\text{RMSE}_{\beta} \leq .031$). Similar to the cluster recovery, the recovery of the regression parameters was mainly influenced by the within-group sample size N_g. When $N_g = 50$, the RMSE_{β} was, on average, 0.059, 0.035, 0.021,

Table 3
Root Mean Squared Error per Level of Each Manipulated Factor for Each of the Four Estimated Regression Parameters

Factor	Level	β_1	β_2	β_3	β_4
β size	.2	.047 (.05)	.037 (.04)	.022 (.04)	.032 (.04)
	.3	.025 (.05)	.010 (.03)	.007 (.02)	.030 (.04)
	.4	.019 (.05)	.002 (.02)	.002 (.02)	.031 (.04)
G	12	.026 (.04)	.012 (.03)	.008 (.02)	.034 (.03)
	24	.030 (.05)	.016 (.04)	.010 (.03)	.031 (.04)
	48	.035 (.06)	.021 (.04)	.014 (.03)	.029 (.04)
N_g	50	.058 (.06)	.035 (.05)	.021 (.04)	.047 (.05)
	100	.021 (.04)	.009 (.03)	.006 (.02)	.027 (.03)
	200	.011 (.02)	.005 (.02)	.004 (.02)	.019 (.02)
K	2	.011 (.02)	.009 (.03)	.000 (.01)	.013 (.01)
	4	.050 (.06)	.023 (.04)	.021 (.04)	.049 (.05)
Cluster size	Balanced	.012 (.02)	.009 (.02)	.006 (.02)	.019 (.02)
	Unbalanced	.048 (.06)	.024 (.05)	.015 (.04)	.043 (.05)
Reliability	High	.030 (.05)	.016 (.03)	.010 (.03)	.031 (.04)
	Low	.030 (.05)	.016 (.03)	.011 (.03)	.031 (.04)
Noninv groups	0.25	.030 (.05)	.016 (.04)	.010 (.03)	.031 (.04)
	0.50	.030 (.05)	.016 (.04)	.011 (.03)	.031 (.04)
	0.75	.030 (.05)	.016 (.04)	.010 (.03)	.031 (.04)
Noninv size	0.4	.030 (.05)	.016 (.04)	.010 (.03)	.031 (.04)
	0.6	.030 (.05)	.016 (.04)	.010 (.03)	.031 (.04)
Total	—	.030 (.05)	.016 (.04)	.010 (.03)	.031 (.04)

Note. β size is the size of (difference in) regression parameters, *G* is the number of groups, N_g is the within-group sample size, *K* is the number of clusters, Noninv *G* is the proportion of noninvariant groups. %CC is the proportion of perfect cluster recovery. CC = correct clustering.

and 0.047 for β_1, β_2, β_3, and β_4, respectively. In contrast, when $N_g = 200$, the RMSE was close to 0 for all regression parameters. From Table 3, it is clear that the parameter recovery of β_1 and β_4 was worse than that of β_2 and β_3. This was consistent across all simulated conditions. Note that β_1 and β_4 both relate to factor *F*4, a purely endogenous factor, whereas β_2 and β_3 both relate to *F*3, a mediating factor. Previous studies reported worse recovery for parameters that were further away from the exogenous latent variables (e.g., Devlieger & Rosseel, 2017; Guenole & Brown, 2014). Indeed, β_4 is further away from the exogenous factors than β_2 and β_3, and the estimation of β_1 depends on that of β_4 since they are both regression effects on *F*4.

Manipulated factors that influenced the recovery for all four regression parameters were the size of the regression coefficients, the number of clusters, and the size of the clusters. Specifically, smaller coefficients, more clusters, and unbalanced clusters resulted in worse recovery. The interaction between the mentioned manipulated factors is shown in Figure 4. Specifically, the largest RMSE values were found when $N_g = 50$, $\beta = .2$, and $K = 4$, and the cluster size is unbalanced. In these conditions, the average RMSE of each regression parameter was close to or above 0.10. Again, increasing the sample size to $N_g = 100$ and/or the size of the regression coefficient to $\beta = .3$ drastically improves the parameter recovery. Lastly, it should be noted that, as long as the sample size is large enough ($N_g = 200$), the RMSE approaches 0 for all combinations of the manipulated factors. Thus, similar trends are found for the RMSE values as for the cluster recovery.

To understand the impact of cluster recovery (i.e., misclassification and classification uncertainties) on the recovery of the regression parameters, we examined the correlation coefficient between the fARI and the RMSE for the four regression parameters, β_1, β_2, β_3, and β_4. The results indicated negative correlations for all four regression parameters, with stronger correlations observed for β_1 (−.70) and β_2 (−.77) than for β_3 (−.56) and β_4 (−.45). These findings suggest that higher RMSE is related to lower fARI and, thus, to more misclassification and/or classification uncertainty. It should be noted that the correlation coefficients were larger for β_1 and β_2 due to the simulation design, in which these parameters were manipulated across all conditions, whereas β_3 and β_4 were only manipulated in conditions with four clusters. In contrast, if we consider only the conditions with four clusters, where all regression parameters were manipulated, the results for β_1 (−.64), β_2 (−.67), β_3 (−.68), and β_4 (−.57) demonstrated a higher degree of similarity.

Simulation Study 2

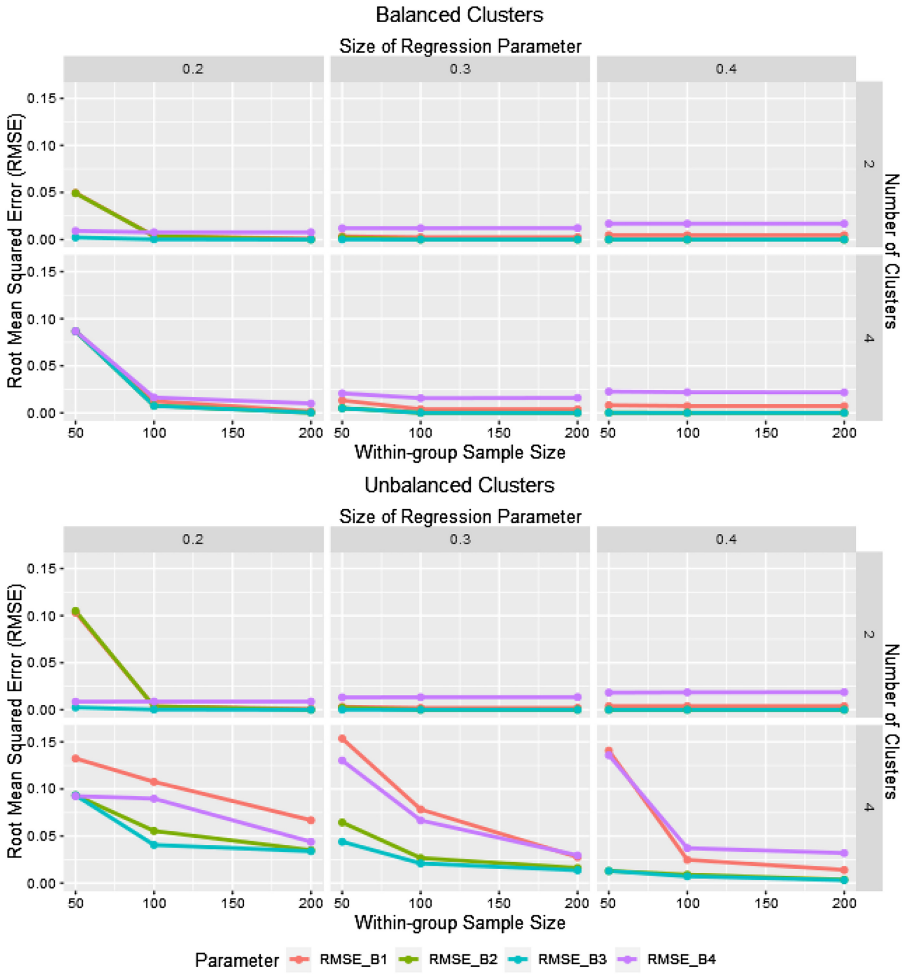
Since the effect of ignoring measurement noninvariances on the estimation of structural relations has been demonstrated in the literature (Chen, 2008; Guenole & Brown, 2014; Pokropek et al., 2019), Simulation Study 2 aimed to investigate how MMG-SEM's performance is affected by disregarding noninvariances. Simulation Study 2 section mostly followed the data generation procedure detailed in Simulation Study 1, but retained only five factors from its design. Furthermore, new factors were added regarding the type of loading noninvariance, the range of residual variance noninvariance, whether these noninvariances were modeled or not, and whether the endogenous factors' residual covariances (EFRCs) were specified as group-and-cluster-specific or cluster-specific (see the Step 2 subsection in the Method). In particular, we manipulated the following factors in Simulation Study 2:

1. Size of regression parameters β (three levels): .2, .3, and .4.
2. Within-group sample size N_g (three levels): 50, 100, and 200.
3. Number of clusters *K* (two levels): 2 and 4.
4. Cluster size (two levels): balanced and unbalanced.
5. Loading noninvariance size (two levels): 0.4 and 0.6.
6. Type of loading noninvariance (two levels): fixed and random.
7. Maximal difference in residual variances (three levels): 0.2 and 0.6.
8. Measurement noninvariance modeled (two levels): yes and no.
9. Specification of EFRCs (two levels): group-and-cluster-specific and cluster-specific.

The number of groups was fixed at 24, and the proportion of noninvariant groups was 0.50. The loadings were initially manipulated according to the low-reliability level from Simulation Study 1 section (i.e., with an average loading of $\sqrt{0.4}$ for all variables other than the marker variable). However, in this study, the group-specific loadings varied around this average loading depending on the "type of loading non-invariance" factor. When the loading noninvariance is "fixed," the loadings were generated according to the procedure described in Simulation Study 1. Thus, we randomly created two equally sized subsets of groups—that is, the invariant and the noninvariant groups—and all noninvariant groups shared the same MM. We expected a subdivision of the groups into subsets with two distinct MMs to have a discernible effect on the clustering when the loading noninvariance is not taken into account. Specifically, if ignoring the noninvariance would cause bias in the regression coefficients, the clustering might (also) capture the two subsets of groups rather than (just) the clustering that is underlying

Figure 4

Regression Parameter Recovery in Function of Within-Group Sample Size, Number of Clusters, Size of Regression Parameters, and Cluster Size



Note. Top: Balanced cluster size conditions. Bottom: Unbalanced cluster size conditions. RMSE = root-mean-squared error. See the online article for the color version of this figure.

the structural relations. In contrast, the "random" noninvariance generation procedure ensured that each loading (other than the ones for the marker variables) was different for each noninvariant group. Firstly, for each noninvariant group, the noninvariance of each loading (i.e., the group-specific deviation from the average loading) was sampled from a uniform distribution with a range of 0.2 centered around the noninvariance size. Specifically, if the loading noninvariance size was 0.6, the deviation was sampled from $U(0.5, 0.7)$ and, in case of a loading noninvariance size of 0.4, it was sampled from $U(0.3, 0.5)$. Secondly, we randomly decided whether the noninvariance was subtracted or added to the original loading (i.e., we added or subtracted the noninvariance to $\sqrt{0.4}$). As a result, the noninvariances were different for all noninvariant groups. Therefore, we expected that ignoring such random loading differences would lead to a different regression bias for each noninvariant group and, thus, make the clustering more difficult to recover in MMG-SEM.

Regarding the residual variances, they were drawn from a uniform distribution with a range that depended on the added "maximal difference in residual variances" factor. Specifically, they were drawn from $U(0.5, 0.7)$ and $U(0.3, 0.9)$ to allow for a maximal difference in the residual variances of 0.2 and 0.6, respectively, where the former distribution is the same as the one used for the "low reliability" conditions in Simulation Study 1. Note that the sampled residual variances still fluctuate around 0.6 in both cases.

The "measurement noninvariance modeled" factor concerned the specification of MMG-SEM when analyzing the data and not the data generation procedure (i.e., it varies within rather than between data sets). In particular, it pertains to whether or not the measurement noninvariances (i.e., the loading and residual variances noninvariances)[8] were captured with group-specific parameters in the estimation of MMG-SEM. When not modeled, the measurement parameters

[8] Note that the manipulation and modeling of the intercept differences are not included in this simulation because the group-centering of the data removes the mean structure, including the item intercepts.

were restricted to be equal across groups (i.e., modeled as invariant), including the noninvariant ones.

Similarly, we were interested in inspecting how misspecifying the EFRC influences the estimation of the regression parameters. Specifically, the "specification of endogenous factors' residual covariances" factor determined whether the EFRC were restricted to be equal within a cluster or were allowed to be group-and-cluser-specific. As the data were generated with group-and-cluster-specific factors' residual variances, forcing them to be equal within a cluster is a model misspecification which might decrease the MMG-SEM's performance. We also found interesting to inspect how MMG-SEM's performance is affected if both the MM and the EFRC are misspecified and how both misspecifications interacted in this respect.

The simulation included 3 (size of regression parameters) × 3 (within-group sample size) × 2 (number of clusters) × 2 (cluster size) × 2 (loading noninvariance size) × 2 (type of loading noninvariance) × 2 (maximal difference in residual variances) = 288 data generation conditions. For each condition, 50 replications were simulated for a total of 14,400 data sets. MMG-SEM was applied four times to each data set, depending on whether the differences in measurement parameters and/or EFRC were properly modeled or not (factors 8 and 9 of the simulation), with the following specifications: (a) noninvariant measurement parameters specified as group-specific and EFRC specified as group-and-cluster-specific, (b) noninvariant measurement parameters restricted to be equal across groups and EFRC specified as group-and-cluster-specific, (c) noninvariant measurement parameters specified as group-specific and EFRC restricted to be equal within a cluster, and (d) measurement parameters restricted to be equal across groups and EFRC restricted to be equal within a cluster. To simplify the description of the results, from now on, these four models will be referred to as Model 1, Model 2, Model 3, and Model 4, respectively.

Results

Cluster Recovery. Table 4 displays the main effects of the manipulated factors on the cluster recovery for the four different MMG-SEM specifications. For each model specification, the manipulated factors showed comparable effects on the cluster recovery as in Simulation Study 1. The cluster recovery improved when the within-group sample size increased, the number of clusters was smaller, and the size of the regression parameters was larger. Overall, the cluster recovery of the correctly specified MMG-SEM model (i.e., Model 1) was good (i.e., mean ARI = 0.89; mean fARI = 0.86), with a perfect recovery (i.e., ARI = 1) for 80% of the data sets. Surprisingly, the model that ignored the measurement noninvariance (i.e., Model 2) unexpectedly demonstrated a comparable cluster recovery (i.e., mean ARI = 0.89; mean fARI = 0.86), indicating a certain degree of robustness of MMG-SEM to misspecified invariant measurement parameters. The loading noninvariance type and size and the maximal differences in the residual variances had minimal impact on the cluster recovery when these measurement noninvariances were disregarded.

The robustness against ignored measurement noninvariances may be explained by the group-and-cluster-specific specification of the EFRC, which leaves room for capturing between-group differences and thus for absorbing the bias when disregarding the measurement noninvariances. Indeed, we observed that ignoring the measurement noninvariances (i.e., Model 2) led to a biased estimation of the group-and-cluster-specific EFRC. This "bias absorption" hypothesis is corroborated further when comparing the results of Models 3 and 4, which misspecified the EFRC as cluster-specific. On average, Model 3, which included the measurement noninvariances, showed an ARI of 0.83 (already slightly worse than that of Model 1). In comparison, Model 4, which disregarded the noninvariances, presented an average ARI of 0.63, which indicates that MMG-SEM lost its robustness against the ignored measurement noninvariances when the EFRCs were erroneously specified as cluster-specific. Furthermore, the size and type of loading noninvariance greatly impacted the cluster recovery of Model 4. Specifically, when the ignored loading noninvariances had a size of 0.4, the mean ARI dropped to 0.71, whereas ignored loading noninvariances of 0.6 led to a mean ARI of 0.55. Similarly, the random loading noninvariances (i.e., ARI 0.52) had a larger effect on Model 4's performance than the fixed noninvariances (i.e., ARI 0.74). In contrast, Model 3 was unaffected by the loading noninvariance size and type and showed an ARI of 0.83 across these conditions. Note that, for all four models, increasing the maximal difference in residual variances had virtually no effect on the cluster recovery.

A more in-depth look at the interaction between the most impactful manipulated factors (i.e., within-group sample size, size of regression parameters, type of loading noninvariance) for each of the four MMG-SEM models is given in Figure 5. It is clear that the performance of Models 1 and 2 is almost identical, irrespective of the manipulated factors. When the EFRCs were misspecified as cluster-specific, ignoring the measurement noninvariances became a relevant issue: Model 4 performed markedly worse than Model 3 in all the conditions represented and even more so in the case of random loading noninvariances.

Regression Parameter Recovery. The overall recovery was good for all regression coefficients and followed the same trend as in Simulation Study 1, that is, the $RMSE_\beta$ of Model 1 was worse for β_1 (i.e., 0.028) and β_4 (i.e., 0.018) compared to β_2 (i.e., 0.016), and β_3 (i.e., 0.010). The same trend was observed for the four models. To simplify the presentation of the results, Table 5 displays the main effects of the manipulated factors on an average RMSE across the four regression parameters. Consistent with expectations, for all four models, the parameter recovery was better when the within-group sample size was higher, and the number of clusters was two. Notably, the RMSE is higher for Model 2 ($RMSE_\beta = .024$) than for Model 1 ($RMSE_\beta = .018$), even though their cluster recovery was virtually the same. Thus, ignoring measurement noninvariances led to a (slightly) larger bias in the regression parameters, despite the lack of effect on the cluster recovery. Comparing the results of Model 3 ($RMSE_\beta = .020$) to Model 4 ($RMSE_\beta = .040$) demonstrates that disregarding measurement noninvariances had a large influence on the regression parameter recovery when the EFRCs were misspecified as cluster-specific, which is at least partially explained by the cluster recovery being worse in that case.

Figure 6 depicts the interaction effect of the manipulated factors most influential on the regression parameters recovery, demonstrating trends similar to those seen for the cluster recovery. When comparing the recovery for Models 1 and 2, it is clear that disregarding the measurement noninvariances did not substantially impact the RMSE when the loading noninvariances were fixed. However, in the case of random noninvariances, the RMSE for Model 2 was higher than that for Model 1, especially for larger regression

Table 4
Average Cluster Recovery Results per Level of Each Manipulated Factor in Simulation Study 2

		Model 1		Model 2		Model 3		Model 4	
Factor	Level	ARI	fARI	ARI	fARI	ARI	fARI	ARI	fARI
β size	.2	0.72 (0.42)	0.65 (0.42)	0.72 (0.42)	0.66 (0.42)	0.64 (0.41)	0.61 (0.39)	0.32 (0.37)	0.32 (0.36)
	.3	0.97 (0.10)	0.96 (0.11)	0.96 (0.11)	0.96 (0.12)	0.93 (0.18)	0.93 (0.16)	0.73 (0.35)	0.74 (0.34)
	.4	0.98 (0.06)	0.98 (0.07)	0.97 (0.11)	0.97 (0.10)	0.93 (0.19)	0.94 (0.17)	0.83 (0.28)	0.84 (0.27)
N_g	50	0.72 (0.42)	0.66 (0.42)	0.72 (0.42)	0.66 (0.41)	0.73 (0.40)	0.70 (0.40)	0.58 (0.42)	0.58 (0.42)
	100	0.96 (0.13)	0.94 (0.15)	0.96 (0.14)	0.94 (0.15)	0.88 (0.24)	0.90 (0.21)	0.65 (0.39)	0.66 (0.38)
	200	0.99 (0.04)	0.99 (0.04)	0.98 (0.09)	0.98 (0.08)	0.89 (0.24)	0.89 (0.22)	0.66 (0.38)	0.66 (0.38)
K	2	0.95 (0.22)	0.91 (0.25)	0.95 (0.22)	0.92 (0.25)	0.95 (0.20)	0.93 (0.22)	0.72 (0.43)	0.72 (0.43)
	4	0.83 (0.32)	0.81 (0.32)	0.82 (0.33)	0.81 (0.33)	0.71 (0.36)	0.73 (0.34)	0.54 (0.35)	0.55 (0.34)
Cluster size	Balanced	0.94 (0.23)	0.91 (0.26)	0.94 (0.23)	0.91 (0.25)	0.92 (0.24)	0.90 (0.26)	0.72 (0.39)	0.71 (0.39)
	Unbalanced	0.83 (0.32)	0.82 (0.32)	0.83 (0.32)	0.81 (0.32)	0.74 (0.35)	0.76 (0.33)	0.54 (0.40)	0.56 (0.39)
Noninv size	0.4	0.89 (0.28)	0.86 (0.30)	0.89 (0.28)	0.86 (0.29)	0.83 (0.31)	0.83 (0.30)	0.71 (0.38)	0.71 (0.37)
	0.6	0.89 (0.28)	0.86 (0.30)	0.88 (0.29)	0.86 (0.30)	0.83 (0.31)	0.83 (0.30)	0.55 (0.41)	0.55 (0.41)
Noninv type	Fixed	0.89 (0.28)	0.86 (0.30)	0.89 (0.28)	0.87 (0.29)	0.83 (0.31)	0.83 (0.30)	0.74 (0.36)	0.74 (0.35)
	Random	0.89 (0.28)	0.86 (0.30)	0.88 (0.29)	0.86 (0.30)	0.83 (0.31)	0.83 (0.31)	0.52 (0.41)	0.53 (0.41)
Res range	0.2	0.89 (0.28)	0.86 (0.30)	0.89 (0.28)	0.86 (0.29)	0.83 (0.31)	0.83 (0.30)	0.64 (0.40)	0.64 (0.39)
	0.6	0.89 (0.28)	0.86 (0.30)	0.88 (0.29)	0.86 (0.30)	0.83 (0.31)	0.83 (0.31)	0.62 (0.40)	0.63 (0.40)
Total		0.89 (0.28)	0.86 (0.30)	0.89 (0.29)	0.86 (0.29)	0.83 (0.31)	0.83 (0.30)	0.63 (0.40)	0.63 (0.40)

Note. Results are shown for the ARI and the fARI. The standard deviation is indicated between brackets. Model 1 includes the MNIs and correctly specifies the EFRCs as group-and-cluster-specific. Model 2 disregards the MNI and correctly specifies the EFRC. Model 3 includes the MNI and misspecifies the EFRC as cluster-specific. Model 4 ignores the MNI and misspecifies the EFRC as cluster-specific. β size is the size of regression parameters, G is the number of groups, N_g is the within-group sample size, and K is the number of clusters. ARI = adjusted rand index; fARI = Frobenious adjusted rand index; Noninv = noninvariance; Res range = maximal difference in residual variances; MNIs = measurement noninvariances; EFRCs = endogenous factors' residual covariances.

parameters. Conversely, when comparing the RMSE for Model 4 to that of Model 3, we see that ignoring the measurement noninvariances resulted in a notably worse recovery for both types of loading noninvariances, and even more so in the case of random loading noninvariances. Lastly, it is interesting to note that increasing the sample size improved the RMSE for models that correctly specify the EFRC (i.e., for Models 1 and 2, the RMSE approaches 0 when $N_g = 200$), where this is not the case for the models that misspecified these EFRC, particularly in conditions where the regression parameters were larger than 0.2.

Conclusion

We assessed the MMG-SEM's performance through two simulation studies. Simulation Study 1 extensively assessed the performance of MMG-SEM when the true number of clusters and the measurement noninvariances were known and correctly specified. We found that performing MMG-SEM with 20 random starts (in Step 2) was sufficient to prevent convergence at local maxima for most data sets, which occurred in only 12.46% of the cases. Nevertheless, since computation time is not an issue, we suggest 50 (or more) starts in more challenging conditions, such as when the within-group sample sizes are small or when a large number of clusters is expected. The recovery of the clustering and regression parameters was unsatisfactory when cluster separation was small (β = .2), the within-group sample size was small ($N_g = 50$), the clusters were unbalanced, and the number of clusters was large ($K = 4$). In these cases, MMG-SEM struggled to recover the clusters, leading to biased estimates of the regression parameters. However, a slight increase in either the within-group sample size or the size of the regression parameters (i.e., $N_g = 100$ or β = .3) significantly improved MMG-SEM's performance, especially in the case of balanced clusters, resulting in an excellent recovery of clusters and regression parameters.

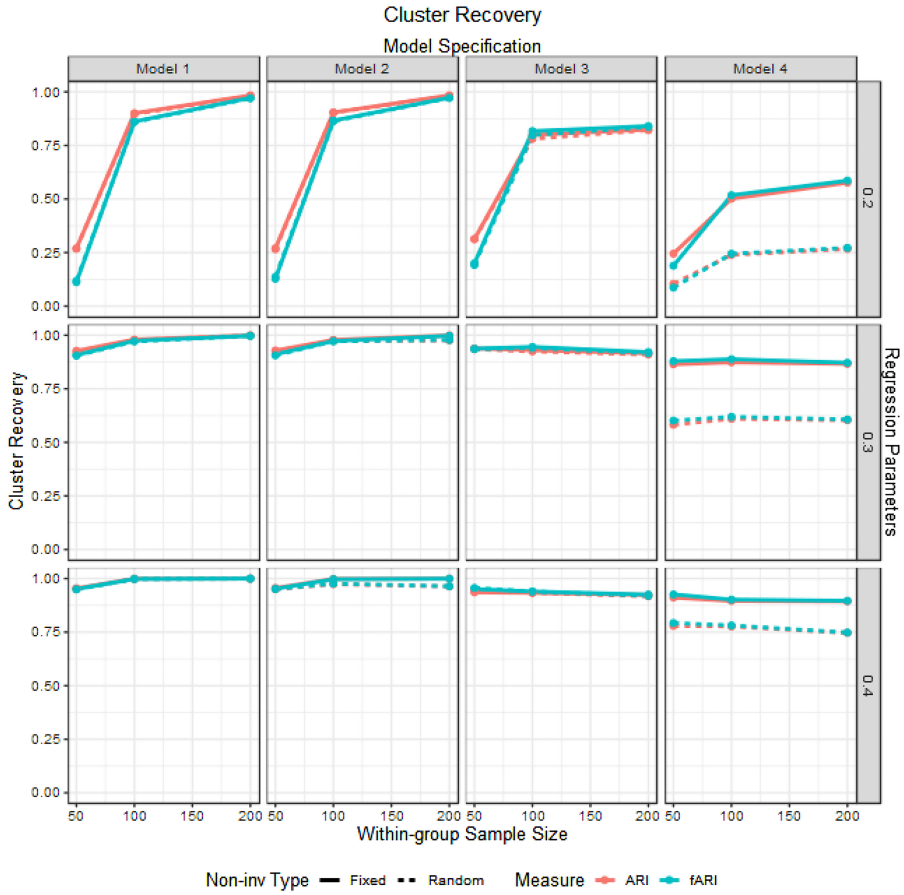
Simulation Study 2 focused on the performance of MMG-SEM when the MM and the EFRC were misspecified. Simulation Study 2 revealed that the performance declined when MMG-SEM disregarded the measurement noninvariances and/or misspecified the EFRC as cluster-specific when they were, in fact, group-specific. Remarkably, the cluster recovery was barely affected by ignoring the measurement noninvariances in the case of group-and-cluster-specific EFRC, indicating some degree of robustness against these noninvariances thanks to the flexibility of the mixture multigroup approach (i.e., the group-and-cluster-specific EFRC absorb some bias, so the clustering was unaffected by it). Ignoring the noninvariances still caused bias in the regression parameters, however, highlighting the importance of assessing the MI beforehand. The robustness against disregarded measurement noninvariance did not hold when the EFRC were incorrectly specified as cluster-specific, a model specification that leaves no room for bias to be absorbed by anything other than the clustering and the corresponding cluster-specific parameters.

Empirical Application

In this section, we demonstrate the empirical value of MMG-SEM by applying it to data collected in the International College Survey of 2001 (ICS; see Kuppens et al., 2006). A questionnaire including self-report items on the frequency of emotions and satisfaction with life was administered to 10,018 participants from 48 different countries. The frequency of emotions was assessed by asking the participants to rate how often they experienced specific emotions during the past week on a 9-point Likert scale (1 = *not at all*, 9 = *all the time*). The measured emotions covered positive (i.e., pleasant, happy, cheerful, pride, gratitude, and love) and negative (i.e., unpleasant, sad, anger, guilt, shame, worry, stress, and jealousy) emotions. Life satisfaction was evaluated with the Satisfaction With Life Scale (SWLS; Pavot

Figure 5

Cluster Recovery in Function of Within-Group Sample Size, Size of the Regression Parameters, Type of Loading Noninvariance, and Model Specification



Note. Non-inv = noninvariance; ARI = adjusted rand index; fARI = Frobenious adjusted rand index. See the online article for the color version of this figure.

& Diener, 1993), which consists of five items rated on a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*).

In previous research, Kuppens et al. (2008) used a multilevel regression model to study how life satisfaction relates to the experience of positive emotions (PEs) and negative emotions (NEs). Specifically, they used the frequency of PE and NE as predictors of life satisfaction, while allowing for random intercepts and slopes at the nation level (i.e., they were allowed to differ across nations). Furthermore, they included individualism and survival/self-expression as moderators to study the impact of national culture on the relation of interest. In this way, the multilevel analysis captured the overall prediction of life satisfaction on experienced emotions and the variation in these relations across the countries. The moderator "individualism" showed a significant effect on the slope of NEs, indicating that differences between countries in the relationship between NEs and life satisfaction depend on individualism. These results suggest that individualistic (as opposed to collectivist) countries emphasize NEs more when making life satisfaction judgments. In contrast, the moderator "survival/self-expression" presented a significant effect on the slope of PEs. Such results suggest that countries that value self-expression allow their citizens to benefit more from positive emotional experiences than countries where their citizens are focused on daily survival. It is important to note, however, that the coefficients of the moderators were small for both individualism (<0.001) and survival/self-expression (0.03) and must be interpreted conservatively.

On the one hand, the multilevel model is inconvenient for pinpointing which specific countries were similar to and/or different from each other regarding the regression coefficients. To compare the individual, country-specific regression effects to one another, one would need 1,081 pairwise comparisons per regression parameter. On the other hand, the multilevel regression model was applied to sum scores of the three variables of interest (i.e., PE, NE, and SWLS). Although using sum scores does not necessarily invalidate an analysis, using them is akin to an overly restrictive latent variable model. McNeish and Wolf (2020) explained that sum scores are equivalent to a factor analysis model where factor loadings and residual variances are constrained to equality for all items of a factor and equal across groups in case of multigroup data (i.e., invariant). This may be an issue considering that not all items are equally reliable and that ignoring potential noninvariances can cause bias in the estimates of the structural parameters (Guenole & Brown, 2014;

Table 5
Average Root Mean Squared Error, Computed Across the Four Regression Parameters, per Level of Each Manipulated Factor in Simulation Study 2

Factor	Level	Model 1	Model 2	Model 3	Model 4
β size	.2	0.034 (0.037)	0.035 (0.036)	0.033 (0.035)	0.050 (0.030)
	.3	0.013 (0.032)	0.019 (0.031)	0.014 (0.031)	0.036 (0.041)
	.4	0.007 (0.025)	0.018 (0.031)	0.012 (0.034)	0.034 (0.048)
N_g	50	0.038 (0.043)	0.041 (0.041)	0.031 (0.039)	0.044 (0.040)
	100	0.012 (0.027)	0.018 (0.028)	0.014 (0.029)	0.037 (0.040)
	200	0.005 (0.018)	0.013 (0.023)	0.015 (0.032)	0.038 (0.042)
K	2	0.005 (0.014)	0.010 (0.014)	0.004 (0.012)	0.021 (0.029)
	4	0.031 (0.042)	0.038 (0.041)	0.036 (0.042)	0.058 (0.042)
Cluster size	Balanced	0.008 (0.020)	0.013 (0.020)	0.008 (0.021)	0.026 (0.030)
	Unbalanced	0.028 (0.041)	0.035 (0.040)	0.032 (0.041)	0.054 (0.045)
Noninv size	0.4	0.018 (0.034)	0.022 (0.033)	0.020 (0.034)	0.032 (0.038)
	0.6	0.018 (0.034)	0.026 (0.034)	0.020 (0.035)	0.048 (0.042)
Noninv type	Fixed	0.018 (0.034)	0.020 (0.033)	0.020 (0.035)	0.028 (0.037)
	Random	0.018 (0.034)	0.028 (0.033)	0.020 (0.035)	0.052 (0.041)
Res range	0.2	0.018 (0.034)	0.024 (0.034)	0.020 (0.035)	0.039 (0.040)
	0.6	0.018 (0.034)	0.024 (0.033)	0.020 (0.035)	0.041 (0.041)
Total		0.018 (0.034)	0.024 (0.034)	0.020 (0.035)	0.040 (0.041)

Note. The standard deviation is indicated between brackets. Model 1 includes the MNIs and correctly specifies theEFRCs as group-and-cluster-specific. Model 2 disregards the MNI and correctly specifies the EFRC. Model 3 includes the MNI and misspecifies the EFRC as cluster-specific. Model 4 ignores the MNI and misspecifies the EFRC as cluster-specific. β size is the size of regression parameters, G is the number of groups, N_g is the within-group sample size, and K is the number of clusters. Noninv = noninvariance; Res range = maximal difference in residual variances; MNIs = measurement noninvariances; EFRCs = endogenous factors' residual covariances.

Pokropek et al., 2019). To resolve these issues, we applied MMG-SEM to the data. The initial SEM model we used is depicted in Figure 7. Given that some countries presented small sample sizes (i.e., $N_g < 50$), we used 50 random starts for Step 2 of MMG-SEM to avoid local maxima (see the Conclusion section). After removing participants with missing data in (one or more of) our variables of interest, we retained a sample of 9,073 participants from 47 countries.[9]

Before applying MMG-SEM, we evaluated MI up to the metric invariance level using MG-CFA for the frequency of emotions in one measurement block and SWLS in another measurement block. We resorted to separate measurement blocks because emotion frequencies and life satisfaction are separate concepts from a substantive point of view—evaluated by separate questionnaires—and, thus, have separate MM. The MI testing revealed residual covariances between some emotion frequency items, some cross-loadings on the emotions factors (e.g., a cross-loading of "pride" on NE factor), and a few noninvariant factor loadings for the emotions factors (i.e., they differed across nations). Specifically, the loading of the item "pride" on the PE was found to be positive for most countries but negative for one nation (i.e., South Africa). Combined with its cross-loading on NE, this suggests that the meaning of pride differs across countries in the sense that, for some countries, pride co-occurs with the experience of NE while being reversely related to PE. Another invariant loading corresponded to the item "shame," which strongly loaded (i.e., $\lambda > 0.80$) on NE for the majority of the countries, but presented relatively lower loadings (i.e., $\lambda < 0.50$) in four nations (i.e., Greece, Switzerland, Singapore, and Kuwait), indicating that it was a weaker indicator of the experience of negative emotions in these countries. Note that the differences in measurement parameters are irrelevant to the research question but can potentially bias the estimations of the regressions that are of interest, as may have happened in the multilevel regression analysis from Kuppens et al. (2008). Since the main focus of MMG-SEM is the clustering of the structural parameters (Step 2), we refer the reader to Appendix B for more details on the MI testing and on the final MG-CFA model used in Step 1 of MMG-SEM.

When analyzing empirical data, we do not know the "true" underlying number of clusters. Therefore, we must select the appropriate number of clusters to use for MMG-SEM. Specifically, we ran Step 2 with one to six clusters and relied on the BIC_G for selecting the best model. Please see Table 6 and Figure 8 for the BIC_G results. The BIC_G indicates that the model with four clusters is the best in terms of fit and parsimony.

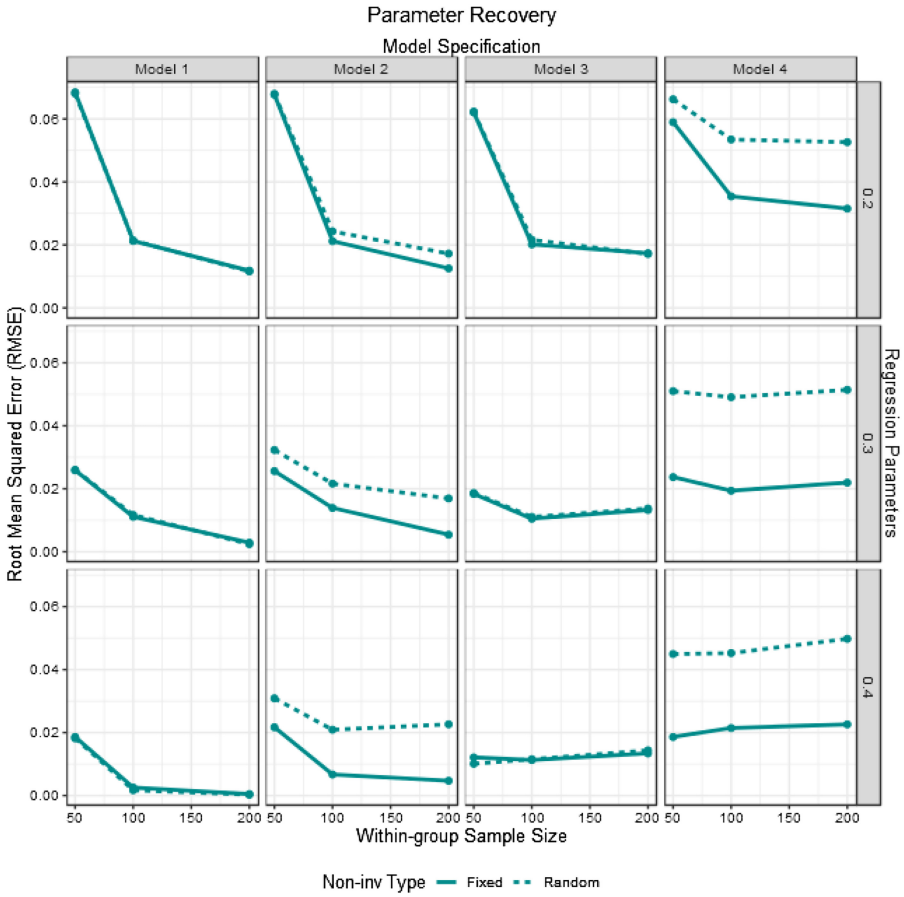
The countries in each one of the resulting clusters can be seen in Table 7. Clearly, there are three small clusters (Clusters 1 to 3) including at most four countries, and one large cluster that contains 36 out of the 47 countries (Cluster 4). Whereas Clusters 1 and 4 contain countries from different geographic locations, Clusters 2 and 3 contain Eastern European (Slovenia and Slovakia) and African (Nigeria and South Africa, next to India) countries, respectively. The cluster-specific regression parameters are listed in Table 8. In the following text, β_{PE} and β_{NE} denote the regression parameter between the frequency of PEs and life satisfaction and between the frequency of NEs and life satisfaction, respectively.

The regression results in Table 8 indicate clear differences between the clusters in how experienced emotions relate to life satisfaction. In Cluster 1, there is a strong positive relation between PEs and life satisfaction ($\beta_{PE} = .44$) and a strong negative relation

[9] Note that Kuppens et al. (2008) removed one country (i.e., Egypt) due to the correlations between the variables of interest being exceptionally high. We also removed it since we want to relate our results to theirs.

Figure 6

Regression Parameter Recovery in Function of Within-Group Sample Size, Size of the Regression Parameters, Type of Noninvariance, and Model Specification



Note. Non-inv = noninvariance. See the online article for the color version of this figure.

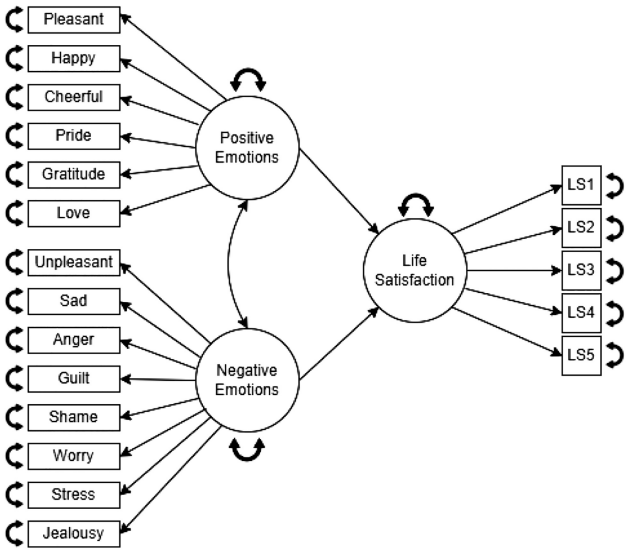
between NEs and life satisfaction ($\beta_{NE} = -.41$), showing that participants from the countries in this cluster have a higher life satisfaction when they experience more PEs and less NEs in the past week. In Cluster 2, both regression parameters are smaller and positive ($\beta_{PE} = .25$; $\beta_{NE} = .08$). That both of them are positive implies that, for the three countries in this cluster, both experiencing PE and NEs predicts a higher life satisfaction, where the relation is slightly stronger for PEs. Cluster 3 presents a small positive β_{PE} (0.15) and a strong negative β_{NE} (−0.59), showing that, for inhabitants of the nations in this cluster, a higher rating of life satisfaction is mostly related to having experienced less NEs in the past week. In contrast, Cluster 4 shows the reverse pattern compared to Cluster 3. It presents a strong positive β_{PE} (0.38) and weaker negative β_{NE} (−0.16), which implies that the life satisfaction of the inhabitants of these countries is predicted more by experiencing PEs than by the lack of negative emotional experiences.

In comparison to the regression coefficients obtained by Kuppens et al. (2008) using a multilevel model (i.e., $\beta_{PE} = .31$ and $\beta_{NE} = -.18$), our parameter estimates were larger for most clusters. This may be attributed to using SEM, which accounts for measurement error and noninvariances that are not considered when applying regression on sum scores. As such, MMG-SEM likely provides a more accurate representation of the true relations between the variables as well as a comprehensive analysis (i.e., capturing differences and similarities in relations).

The relationship between emotions and life satisfaction, including cross-cultural differences, has been extensively examined in previous research (e.g., Bastian et al., 2014; Cohn et al., 2009; Kang et al., 2003; Kuppens et al., 2006; Zou et al., 2018). Notably, indications were found that individualistic countries have a slightly stronger (negative) association between NE and life satisfaction than collectivistic countries (Kuppens et al., 2008; Suh et al., 1998). These findings can be attributed to cultural differences in emotional regulation strategies (Bastian et al., 2012; Miyamoto et al., 2014). To investigate whether the regression parameters from MMG-SEM depended on individualism, we inspected the level of individualism from Hofstede (2001) for different nations, which was also used by Kuppens et al. (2008), and was available for 41 of the 47 countries from the ICS. The individualism score ranged from 0 to 100, with higher scores indicating a more individualistic culture. The left-hand side of Figure 9 depicts the individualism results for the four clusters from MMG-SEM. This figure indicates that, among the three smaller

Figure 7

The SEM Model Depicting the Relations Between Emotions and Life Satisfaction Evaluated for the Data From the International College Survey 2001



Note. The indicators of the factor "life satisfaction" are described in Pavot and Diener (1993). SEM = structural equation modeling; LS = life satisfaction.

clusters, Cluster 1 contained the most individualistic countries, whereas Cluster 3 collected the least individualistic countries. The individualism scores varied considerably within each cluster, however. For instance, Cluster 4, the largest cluster, included countries with scores ranging from 12 (Venezuela) to 90 (Australia). Moreover, the individualism scores overlapped substantially across the four clusters, suggesting that individualism may not be the (only) variable affecting the relationship between emotions and life satisfaction.

Following Kuppens et al. (2008) research, we also inspected the impact of "survival/self-expression" (S/SE) from Inglehart and Baker (2000) on the relation between emotions and life satisfaction. The "survival/self-expression" dimension measures whether the citizens of a nation focus on survival (i.e., prioritize economic and physical security, as indicated by lower scores) or self-expression (i.e., prioritize the quality of life and self-expression, as indicated by higher scores). Scores were available for 39 out of the 47 countries from the ICS. The boxplots of the S/SE scores for the four clusters are presented on the right-hand side of Figure 9. Within the three smaller clusters, Cluster 1 exhibited the highest S/SE scores, while Clusters 2 and 3 comprised countries with lower S/SE values. Although these findings align with Kuppens et al. (2008) (i.e., a higher β_{PE} is associated with higher S/SE scores), it is noteworthy that Cluster 4 contained countries with both the highest (i.e., the Netherlands) and lowest S/SE scores (i.e., Russia) of the whole sample. This suggests that, at least for Cluster 4, S/SE is not the only variable influencing the relationship between emotions and life satisfaction.

Table 6

BIC_G *for the Models From 1 to 6 Clusters*

Clusters	BIC_G	LogL
1	649,631.55	−321,161.99
2	649,629.34	−321,155.10
3	649,630.83	−321,150.07
4	**649, 621.06**	−321,139.41
5	649,631.53	−321,138.88
6	649,643.21	−321,138.94

Note. LogL is the log-likelihood of the observed data (Equation 9). BIC = Bayesian information criterion. The best model (i.e., lower BIC) can be seen in bold.

In summary, MMG-SEM unveiled intriguing cross-national differences in the relation between experienced emotions and life satisfaction. We identified four distinct clusters with meaningful differences in the structural relations of interest. Compared to the multilevel regression analysis by Kuppens et al. (2008), MMG-SEM allowed us to avoid a large number of pairwise comparisons and obtain an insightful overview of which countries have comparable structural relations. Furthermore, we were able to easily examine whether individualism and S/SE differed across the clusters of countries by visually inspecting boxplots. Our results demonstrated minimal differences in individualism across the four clusters and, thus, that individualism seemingly does not explain the between-cluster differences in relations. It should be noted that these results do not diverge from the conclusions of Suh et al. (1998) and Kuppens et al. (2008), since although statistically significant, the moderator effect of individualism by Kuppens et al. (2008) was extremely small, indicating that the differences in relations between individualistic and collectivist cultures were minor. In contrast, in our S/SE results, Clusters 1 to 3 presented interesting differences. Cluster 1 presented a clearly higher S/SE score than Clusters 2 and 3, which is in line with Kuppens et al. (2008). Nevertheless, for both individualism and S/SE, Cluster 4 did not show any discernible pattern and contained countries with a wide range of scores. It is likely that other variables also play a role in shaping the relationship between emotions and life satisfaction. For instance, Kang et al. (2003) proposed a model where "emotional expression" and "emotional differentiation" can influence how emotions relate to life satisfaction.

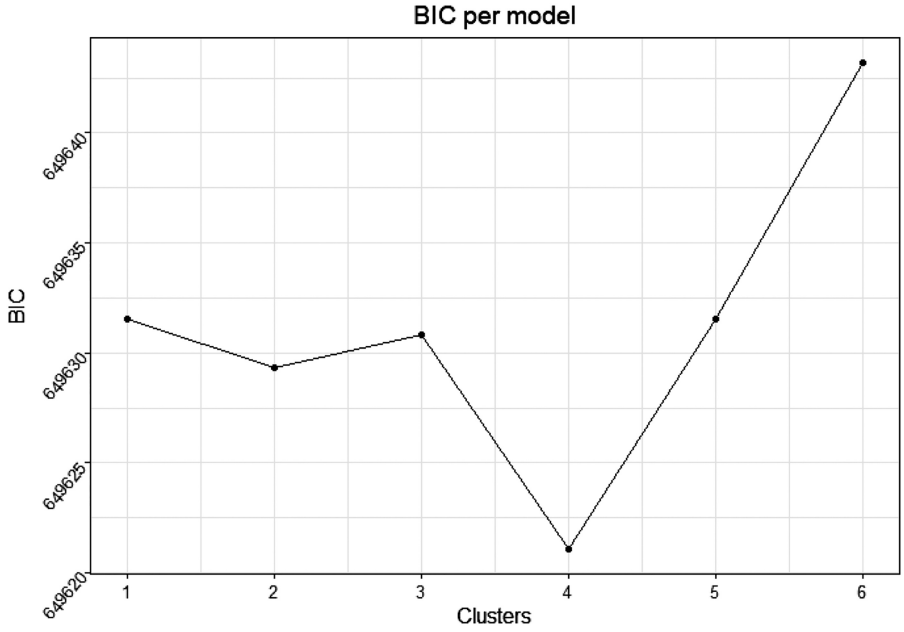
Discussion

This article presented MMG-SEM as a novel method to validly and efficiently compare structural relations (i.e., regressions between latent variables) among many groups. Unlike other mixture SEM methods, MMG-SEM's flexibility allows disentangling differences in the structural relations (of interest to the research question) from other sources of differences (e.g., in the MM). By using a combination of group- and/or cluster-specific parameters (i.e., the mixture multigroup approach), it accounts for potential measurement noninvariances, while not letting them determine the clustering. In an elaborate simulation study, MMG-SEM showed promising results as it correctly identified clusters of groups based only on their structural relations. By assigning the many groups to a few clusters, it efficiently captures differences and similarities in the structural relations. For instance, in an empirical example, MMG-SEM was applied to data previously analyzed by Kuppens et al. (2008) to compare the relations between emotions and life satisfaction across 47 countries, resulting in four clusters of countries that strongly differed in these relations.

While clustering simplifies the comparison among many groups, it also raises the issue of selecting the number of clusters. For the empirical application, we relied on the commonly used BIC (Schwarz, 1978), which had already demonstrated a good performance for

Figure 8

Bayesian Information Criterion (BIC) Results for Models From One to Six Clusters



existing mixture multigroup methods (De Roover, 2021; De Roover et al., 2022). However, other methods for model selection are available. For instance, the CHull method was proven to be a valuable alternative to BIC (Bulteel et al., 2013; De Roover, 2021; De Roover et al., 2022). In addition to methods that balance model fit and parsimony, methods that give more importance to substantive interpretation are also of interest (e.g., latent class trees; van den Bergh et al., 2017). If small differences in structural relations are not relevant to the research question, one should be able to consider this when selecting the model complexity. In future research, we will extensively evaluate the merits and pitfalls of different model selection methods for selecting the number of clusters in MMG-SEM.

To account for measurement noninvariances when clustering groups on their structural relations, MMG-SEM assumes that an MI test has been performed beforehand. In this article, we used MG-CFA for estimating the MM and evaluating its invariance. However, the traditional MI testing using MG-CFA (see Vandenberg & Lance, 2000) can be inefficient when many groups are involved. Many alternative methods to evaluate invariance are available (see Leitgöb et al., 2022). Thanks to SAM's step-wise estimation, one can easily replace MG-CFA with one of these alternatives in Step 1 while Step 2 remains the same. For instance, one could use the approximate invariance (Muthen & Asparouhov, 2013) approach from the Bayesian SEM framework, which relaxes the requirement of exactly invariant parameters. It allows for small differences in measurement parameters (e.g., small variations in loadings across groups), which is often more realistic for applications with many groups. Similarly, one could apply multilevel factor analysis (see Kim et al., 2017; Leitgöb et al., 2022), which treats the groups as random instead of fixed and is, thus, especially suitable for many groups. Specifically, one can add random effects to capture differences in the measurement parameters across groups.

Clearly, the SAM step-wise estimation has demonstrated many benefits. In this article, we only applied the local variant of SAM, however. Even though Rosseel and Loh (2022) stated that local and global SAMs are equivalent for standard SEM, this equivalence has not been tested in the context of mixture SEM models. Furthermore, we chose a step-wise approach due to its advantages but did not compare its performance to the one-step estimation. To explore the strengths and weaknesses of, and differences between one-step and (local and

Table 7

Clustering of the Countries Based on the Regression Parameters Between Emotions and Life Satisfaction (Four-Cluster Model)

Model	Countries
Cluster 1	United States, Japan, Portugal, and Kuwait
Cluster 2	Nigeria, Greece (0.77), and India, South Africa (0.84).
Cluster 3	Slovenia, Uganda, Ghana, and Slovakia
Cluster 4	Turkey, Korea, Rep., Chile, China, Thailand, Australia (0.98), Hong Kong, Iran (0.98), Hungary, Philippines, Poland, Nepal, Spain, Cyprus (0.98), Switzerland (0.97), Indonesia, Italy, Mexico, Belgium (0.97), Venezuela, Singapore (0.94), Netherlands (0.93), Malaysia, Georgia, Croatia, Bulgaria, Bangladesh, Russia, Zimbabwe (0.97), Germany, Colombia, Brazil, Cameroon (0.96), and Canada, Austria (0.82)

Note. Most countries presented a posterior probability (i.e., cluster membership) $\hat{z}_{gk} > 0.99$. The countries with lower $\hat{z}_{gk}$ are presented along with their corresponding posterior probabilities, enclosed in parentheses.

Table 8

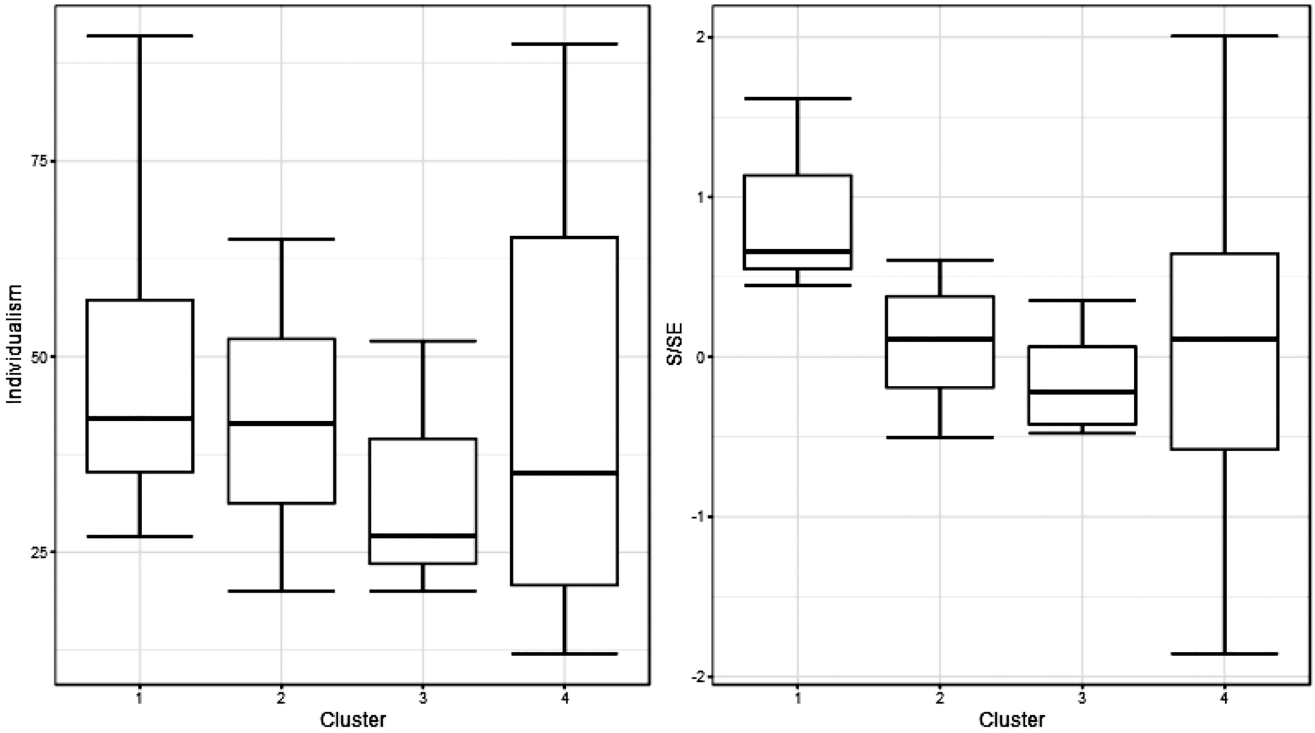
Regression Parameters per Cluster for the Relation Between Emotions and Life Satisfaction

Parameter	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Life satisfaction ~ positive emotions (β_{PE})	.4466	.2526	.1494	.3851
Life satisfaction ~ negative emotions (β_{NE})	−.4182	.0848	−.5906	−.1684

Note. PEs = positive emotions; NEs = negative emotions.

Figure 9

Boxplots for Individualism (Left) and S/SE (Right) Scores for the Four Clusters Obtained by MMG-SEM



Note. S/SE = survival/self-expression; MMG-SEM = mixture multigroup structural equation modeling. The lines for each bar in the boxplot represent the quartiles in the data (i.e., minimum, first quartile, median, third quartile, maximum).

global) two-step estimation approaches, we plan to evaluate the performance of MMG-SEM with different estimation methods.

In the Simulation Study 2, we tested the effect of ignoring measurement noninvariance (i.e., constraining noninvariant parameters to be equal across groups) on the recovery of the clustering and regression parameters. Surprisingly, we found that MMG-SEM demonstrated robustness against ignoring measurement noninvariances in terms of cluster recovery. Despite the unaffected cluster recovery, some bias of the regression parameters (i.e., structural relations) occurred when disregarding measurement noninvariance, indicating that an accurate evaluation of MI remains important. The robustness with regard to the cluster recovery is explained by the fact that allowing for group-and-cluster-specific EFRC largely absorbed the bias, which further underlines the merit of the mixture multigroup approach. To confirm this, we inspected the effect of ignoring measurement noninvariance when incorrectly forcing the EFRC to be equal within clusters. Indeed, when the EFRCs were misspecified as such, ignoring measurement noninvariances led to substantial bias in estimating the regression parameters. The mentioned robustness can also be partly attributed to the manner in which the noninvariances were applied in the simulation study. While we added the noninvariances to all latent variables simultaneously, Chen (2008) suggested that doing so can cause the effect of disregarding the noninvariances to cancel out.

In future research, it would be interesting to also test the robustness of MMG-SEM in case of other common misspecifications. One of the main advantages of the SAM approach is its robustness against local model misspecifications (i.e., omitted cross-loadings and/or residual covariance between items; Rosseel & Loh, 2022). Note that Robitzsch (2022) showed that the SAM step-wise estimation is generally not more (or less) robust than other SEM estimation approaches in case of misspecifications that are spread all over the model (i.e., global misspecifications). Therefore, when testing the performance of MMG-SEM with different estimation approaches (both one-step and step-wise), we will also evaluate their robustness against misspecifications.

MMG-SEM can be extended in several ways. Firstly, MMG-SEM currently uses ML estimation and, thus, assumes the indicator variables to be continuous and normally distributed. However, many studies in social sciences, including the empirical application in this article, use questionnaires with a Likert scale and often show deviations from normality (i.e., skewness and kurtosis; Cain et al., 2017). In the case of nonsevere nonnormality, having five or more response categories allows for considering the ordinal scale as continuous (Johnson & Creech, 1983; Norman, 2010). To accommodate data with severe nonnormality and/or less than five categories, in future research, we plan to include robust estimators for nonnormal data (e.g., Mîndrilă, 2010) and ordinal estimators for noncontinuous data (Agresti, 2010).

A second extension relates to the target parameters of the clustering. As presented in this article, MMG-SEM captures differences in regression parameters among many groups by modeling the covariance structure of the data (i.e., group-specific covariances). Alternatively, one can include the mean structure of the data. It then becomes possible to also cluster groups based on their mean scores on the latent variables (i.e., factor means) if this is relevant to the research question. Note that this

implies that a higher level of MI should hold (i.e., scalar invariance), at least partially.

A third possible extension, is applying the principles of MMG-SEM to longitudinal data where each person was measured at several time points. SEM is also commonly used in longitudinal data to study how latent variables affect each other across time (i.e., dynamic SEM; Asparouhov et al., 2018). To this aim, MMG-SEM can be extended to treat each person as a group and to capture clusters of subjects that share similar dynamic structural relations over time.

To conclude, MMG-SEM is a promising method that allows us to efficiently compare structural relations among many groups while accounting for potential measurement noninvariances. Thanks to its flexibility and its potential for further extensions, it opens up many possibilities to examine differences and similarities that pertain to the latent variable of interest, and that are thus at the heart of the research question.

References

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Wiley.

Andrews, J. L., Browne, R., & Hvingelby, C. D. (2022). On assessments of agreement between fuzzy partitions. *Journal of Classification*, *39*(2), 326–342. https://doi.org/10.1007/s00357-021-09407-3

Arminger, G., & Stein, P. (1997). Finite mixtures of covariance structure models with regressors: Loglikelihood function, minimum distance estimation, fit indices, and a complex example. *Sociological Methods & Research*, *26*(2), 148–182. https://doi.org/10.1177/0049124197026002002

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Bastian, B., Kuppens, P., De Roover, K., & Diener, E. (2014). Is valuing positive emotion associated with life satisfaction? *Emotion*, *14*(4), 639–645. https://doi.org/10.1037/a0036466

Bastian, B., Kuppens, P., Hornsey, M. J., Park, J., Koval, P., & Uchida, Y. (2012). Feeling bad about being sad: The role of social expectancies in amplifying negative mood. *Emotion*, *12*(1), 69–80. https://doi.org/10.1037/a0024755

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.

Brandt, M. J., Turner-Zwinkels, F. M., Karapirinler, B., Van Leeuwen, F., Bender, M., van Osch, Y., & Adams, B. (2021). The association between threat and politics depends on the type of threat, the political domain, and the country. *Personality and Social Psychology Bulletin*, *47*(2), 324–343. https://doi.org/10.1177/0146167220946187

Bulteel, K., Wilderjans, T. F., Tuerlinckx, F., & Ceulemans, E. (2013). CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers. *Behavior Research Methods*, *45*(3), 782–791. https://doi.org/10.3758/s13428-012-0293-y

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Cain, M. K., Zhang, Z., & Yuan, K. -H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, *49*(5), 1716–1735. https://doi.org/10.3758/s13428-016-0814-1

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cohn, M. A., Fredrickson, B. L., Brown, S. L., Mikels, J. A., & Conway, A. M. (2009). Happiness unpacked: Positive emotions increase life satisfaction by building resilience. *Emotion*, *9*(3), 361–368. https://doi.org/10.1037/a0015952

Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., & Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology*, *43*(4), 558–575. https://doi.org/10.1177/0022022112438397

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *39*(1), 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

De Roover, K. (2021). Finding clusters of groups with measurement invariance: Unraveling intercept non-invariance with mixture multigroup factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(5), 663–683. https://doi.org/10.1080/10705511.2020.1866577

De Roover, K., Vermunt, J. K., & Ceulemans, E. (2022). Mixture multigroup factor analysis for unraveling factor loading noninvariance across many groups. *Psychological Methods*, *27*(3), 281–306. https://doi.org/10.1037/met0000355

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM? *Methodology*, *13*(Supplement 1), 31–38. https://doi.org/10.1027/1614-2241/a000130

Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in Psychology*, *5*, Article 980. https://doi.org/10.3389/fpsyg.2014.00980

Hofstede, G. (2001). *Culture's consequences comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Sage Publications.

Hoyle, R. H. (2023). *Handbook of structural equation modeling* (2nd ed.). The Guilford Press.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218. https://doi.org/10.1007/BF01908075

Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, *65*(1), 19–51. https://doi.org/10.1177/000312240006500103

Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, *48*(3), 398–407. https://doi.org/10.2307/2095231

Kang, S.-M., Shaver, P. R., Sue, S., Min, K.-H., & Jing, H. (2003). Culture-specific patterns in the prediction of life satisfaction: Roles of emotion, relationship quality, and self-esteem. *Personality and Social Psychology Bulletin*, *29*(12), 1596–1608. https://doi.org/10.1177/0146167203255986

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 524–544. https://doi.org/10.1080/10705511.2017.1304822

Kim, E. S., Joo, S. -H., Lee, P., Wang, Y., & Stark, S. (2016). Measurement invariance testing across between-level latent classes using multilevel factor mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(6), 870–887. https://doi.org/10.1080/10705511.2016.1196108

Kuppens, P., Ceulemans, E., Timmerman, M. E., Diener, E., & Kim-Prieto, C. (2006). Universal intracultural and intercultural dimensions of the recalled frequency of emotional experience. *Journal of Cross-Cultural Psychology*, *37*(5), 491–515. https://doi.org/10.1177/0022022106290474

Kuppens, P., Realo, A., & Diener, E. (2008). The role of positive and negative emotions in life satisfaction judgment across nations. *Journal of Personality and Social Psychology*, *95*(1), 66–75. https://doi.org/10.1037/0022-3514.95.1.66

Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & van de Schoot, R. (2022). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art,

and future perspectives. *Social Science Research*, *110*, Article 102805. https://doi.org/10.1016/j.ssresearch.2022.102805

Lukočienė, O., & Vermunt, J. K. (2009). Determining the number of components in mixture models for hierarchical data. In A. Fink, B. Lausen, W. Seidel, & A. Ultsch (Eds.), *Advances in data analysis, data handling and business intelligence* (pp. 241–249). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-01044-6_22

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application*, *6*(1), 355–378. https://doi.org/10.1146/annurev-statistics-031017-100325

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

McNicholas, P. D. (2017). *Mixture model-based classification*. CRC Press, Taylor & Francis Group.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Meuleman, B. (2019). Multilevel structural equation modeling for cross-national comparative research. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *71*(S1), 129–155. https://doi.org/10.1007/s11577-019-00605-x

Miyamoto, Y., Ma, X., & Petermann, A. G. (2014). Cultural differences in hedonic emotion regulation after a negative event. *Emotion*, *14*(4), 804–815. https://doi.org/10.1037/a0036257

Mîndrilă, D. (2010). Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal for Digital Society*, *1*(1), 60–66. https://doi.org/10.20533/ijds.2040.2570.2010.0010

Muthen, B., & Asparouhov, T. (2013). *BSEM measurement invariance analysis*. Mplus Web Notes: No. 17. https://www.statmodel.com/examples/webnotes/webnote17.pdf

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Pavot, W., & Diener, E. (1993). Review of the satisfaction with life scale. *Psychological Assessment*, *5*(2), 164–172. https://doi.org/10.1037/1040-3590.5.2.164

Pokropek, A., Davidov, E., & Schmidt, P. (2019). A Monte Carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(5), 724–744. https://doi.org/10.1080/10705511.2018.1561293

Rai, D., Zitko, P., Jones, K., Lynch, J., & Araya, R. (2013). Country- and individual-level socioeconomic determinants of depression: Multilevel cross-national comparison. *British Journal of Psychiatry*, *202*(3), 195–203. https://doi.org/10.1192/bjp.bp.112.112482

R Core Team. (2022). *R: A language and environment for statistical computing*. https://www.R-project.org/

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*(3), 552–566. https://doi.org/10.1037/0033-2909.114.3.552

Robitzsch, A. (2022). Comparing the robustness of the structural after measurement (SAM) approach to structural equation modeling (SEM) against local model misspecifications with alternative estimation approaches. *Stats*, *5*(3), 631–672. https://doi.org/10.3390/stats5030039

Rosseel, Y. (2012). lavaan: An *R* package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y., & Loh, W. W. (2022). A structural after measurement approach to structural equation modeling. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000503

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, *74*(1), 31–57. https://doi.org/10.1177/0013164413498257

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Shireman, E. M., Steinley, D., & Brusco, M. J. (2016). Local optima in mixture modeling. *Multivariate Behavioral Research*, *51*(4), 466–481. https://doi.org/10.1080/00273171.2016.1160359

Steenkamp, J., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–107. https://doi.org/10.1086/209528

Steinley, D. (2006). Profiling local optima in K-means clustering: Developing a diagnostic technique. *Psychological Methods*, *11*(2), 178–192. https://doi.org/10.1037/1082-989X.11.2.178

Steinley, D., & Brusco, M. J. (2011). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods*, *16*(1), 63–79. https://doi.org/10.1037/a0022673

Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology*, *9*(1), 1–12. https://doi.org/10.1027/1614-2241/a000049

Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology*, *74*(2), 482–493. https://doi.org/10.1037/0022-3514.74.2.482

Tueller, S. J., Drotar, S., & Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(1), 110–131. https://doi.org/10.1080/10705511.2011.534695

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

van den Bergh, M., Schmittmann, V. D., & Vermunt, J. K. (2017). Building latent class trees, with an application to a study of social capital. *Methodology*, *13*(Supplement 1), 13–22. https://doi.org/10.1027/1614-2241/a000128

Varriale, R., & Vermunt, J. K. (2012). Multilevel mixture factor models. *Multivariate Behavioral Research*, *47*(2), 247–275. https://doi.org/10.1080/00273171.2012.658337

Vermunt, J. K., & Magidson, J. (2005). Structural equation modeling: Mixture models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons. https://doi.org/10.1002/0470013192.bsa600

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*(298), 348–368. https://doi.org/10.1080/01621459.1962.10480664

Zercher, F., Schmidt, P., Cieciuch, J., & Davidov, E. (2015). The comparability of the universalism value over time and across countries in the European Social Survey: Exact vs. approximate measurement invariance. *Frontiers in Psychology*, *6*, Article 733. https://doi.org/10.3389/fpsyg.2015.00733

Zou, T., Su, Y., & Wang, Y. (2018). Examining relationships between social capital, emotion experience and life satisfaction for sustainable community. *Sustainability*, *10*(8), Article 2651. https://doi.org/10.3390/su10082651

(*Appendices follow*)

Appendix A

Algorithm

The algorithm of MMG-SEM is divided into two steps according to the local variant of the SAM approach from Rosseel and Loh (2022). Below, a detailed description of both steps is presented. The first step is focused on the estimation of the MM, whereas the second step estimates the SM with a mixture clustering of the groups.

Step 1: MM With Measurement Noninvariances

In Step 1, an MG-CFA is estimated in lavaan, using the marker-variable approach to scale the latent variables. When the user specifies more than one measurement block (see Rosseel & Loh, 2022), the MG-CFA is done for each block. Note that the observed scores for each item are centered per group to remove the mean structure of the data. This is equivalent to having group-specific intercepts and factor means. The centering allows us to improve the computational efficiency of the algorithm by omitting their estimation. The default setting regarding the other measurement parameter are the following:

1. The loadings $\mathbf{\Lambda}$ are invariant across all groups, unless the user specifies certain loadings to be noninvariant. User-specified noninvariant loadings are estimated as group-specific. By default, the first nonzero loading of each factor is set to 1 to set the scale of the latent variable (marker indicator approach).
2. The unique variances on the diagonal of $\mathbf{\Theta_g}$ are group-specific, unless the user specifies certain unique variances to be invariant. If $\mathbf{\Theta_g}$ contains residual item covariances, they are also group-specific, unless otherwise specified.
3. The intercepts $\boldsymbol{\tau}_g$ are group-specific and zero due to the centering of the data per group, unless otherwise specified.

Although these settings are applied by default, one should conduct an MI test before applying MMG-SEM, and include all identified noninvariances in the measurement part of the model. If the user needs to deviate from our default specification, they can do so by modifying the same lavaan syntax for Step 1. If invariant intercepts are specified, the group-centering of the data is skipped, and the intercepts (and factor means) are included in the model.

If there is one measurement block, the resulting group-specific factor covariance matrices $\mathbf{\Phi}_g^{s1}$ are the input of Step 2. In case of multiple measurement blocks, group-specific "mapping" matrices $\mathbf{M}_g$ (Rosseel & Loh, 2022) are used to obtain the factor covariance matrices $\mathbf{\Phi}_g^{s1}$. The "mapping" matrix is computed as $\mathbf{M}_g = (\mathbf{\Lambda}^T\mathbf{\Theta}_g^{-1}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^T\mathbf{\Theta}_g^{-1}$, where the $\mathbf{\Lambda}$ and $\mathbf{\Theta}_g$ matrices are assembled using the MM parameters from each measurement block (as shown in the example by Rosseel & Loh, 2022, Appendix G). The factor covariance matrices are then computed as $\mathbf{\Phi}_g^{s1} = \mathbf{M}_g(\mathbf{S}_g - \mathbf{\Theta}_g)\mathbf{M}_g^T$, where $\mathbf{S}_g$ is the observed covariance matrix of the items in group g.

Step 2: SM With Mixture Clustering of the Groups

In Step 2, an EM algorithm (Dempster et al., 1977) is used to cluster the groups based on the structural relations by means of a mixture model. The EM algorithm alternates between two steps: the expectation (E) step and the maximization (M) step. During the E-step, the algorithm estimates the cluster memberships of all groups, resulting in a $G \times K$ matrix of estimated posterior classification probabilities $\mathbf{Z}$, where each element $\hat{\mathbf{z}}_{gk}$ represents the probability that group g belongs to cluster k. Each posterior probability $\hat{\mathbf{z}}_{gk}$ is the estimate or the expected value of the true (binary) cluster membership $\mathbf{z}_{gk}$, formally defined as follows:

$$\hat{\mathbf{z}}_{gk} = E[\mathbf{z}_{gk} = 1|\mathbf{H}_g] = \frac{\boldsymbol{\pi}_k f_{gk}(\mathbf{H}_g;\, \vartheta_{gk})}{\sum_{k=1}^{K} \boldsymbol{\pi}_k f_{gk}(\mathbf{H}_g;\, \vartheta_{gk})}, \tag{A1}$$

where π_k is the prior probability of a group g belonging to cluster k, f_{gk} is the density function of the group g in the kth cluster, and ϑ_{gk} is its corresponding set of parameters. Then, the M-step maximizes all sets of parameters ϑ_{gk} given the cluster memberships $\hat{\mathbf{z}}_{gk}$ obtained in the E-step. For MMG-SEM, this implies estimating the SM parameters for all group-cluster combinations. A detailed step-by-step description of Step 2 reads as follows:

1. Initialize $\mathbf{Z}$ by means of a random hard partition (i.e., with 0/1 cluster memberships) where all clusters contain at least one group.
2. Iterate the following steps while $(\text{LogL}_{\boldsymbol{\eta}_i} - \text{LogL}_{\boldsymbol{\eta}_{i-1}}) > 1 \times 10^{-6}$ or $i \leq 10,000$:
   a. Update the iteration $i = i + 1$.
   b. Update the mixing proportions π_k for all $k = 1, \ldots, K$ clusters by using:

$$\pi_k = \frac{1}{G}\sum_{g=1}^{G} \mathbf{z}_{gk}. \tag{A2}$$

   c. Per cluster $k = 1, \ldots, K$, compute a weighted average of the $\mathbf{\Phi}_g^{s1}$ from step 1. Specifically, this weighted average factor covariance matrix $\mathbf{\Phi}_k^{s1}$ is computed as follows:

$$\mathbf{\Phi}_k^{s1} = \sum_{g=1}^{G} \mathbf{\Phi}_g^{s1} w_{gk}, \tag{A3}$$

   where the weights w_{gk} are determined as $w_{gk} = \frac{\mathbf{z}_{gk} N_g}{N}$, for which N_g is the within-group sample size, and N is the total sample size.[10]

   Note that if the user requested group-and-cluster-specific endogenous factor residual covariances $\mathbf{\Psi}_{gk}^{\text{endog1}}$ and $\mathbf{\Psi}_{gk}^{\text{endog2}}$, a bias correction procedure is performed in this and the following steps, to prevent a biased $\mathbf{B}_k$ in the next iterations ($i > 1$). This correction is described in the next subsection.

   d. Estimate the SM per cluster by calling lavaan using $\mathbf{\Phi}_k^{s1}$ as input. The starting values for each iteration are defined using the ordinary least squares (OLS) estimates, which is the default method employed by lavaan.[11] Since the input $\mathbf{\Phi}_k^{s1}$ varies across iterations, due to updates of the clustering,

[10] Note that this is done per cluster to improve the speed of the algorithm. Using lavaan to estimate a SM for each group-cluster combination is not computationally efficient due to the large number of parameters involved in all combinations.

[11] Note that using the OLS-based starting values accelerates the convergence of lavaan, resulting in only one iteration in most cases.

(*Appendices continue*)

the OLS starting values change correspondingly. Extract the resulting cluster-specific regression parameters $\mathbf{B}_k$ and factor residual covariance matrices $\mathbf{\Psi}_k$.

e. Reconstruct the group-(and-cluster)-specific parts of the group-and-cluster-specific residual factor covariance matrix $\mathbf{\Psi}_{gk}$ (see Equation 7) by using $\mathbf{\Phi}_g^{s1}$ and $\mathbf{\Psi}_k$ as follows:
   i. The group-specific exogenous factor covariances are directly taken from $\mathbf{\Phi}_g^{s1}$.
   ii. The group-and-cluster-specific endogenous factor residual covariances are computed by solving Equation 4 for $\mathbf{\Psi}_{gk}$. That is, we extract the residual covariances from $\mathbf{\Psi}_{gk} = (\mathbf{I} - \mathbf{B}_k)^{-1}\mathbf{\Phi}_g^{s1}(\mathbf{I} - \mathbf{B}_k)^{-1'}$. If the user requested cluster-specific rather than group-and-cluster-specific endogenous residual factor covariances (such that they also drive the clustering), they correspond to the ones in $\mathbf{\Psi}_k$.

f. Compute the $\text{LogL}_{\boldsymbol{\eta}_i}$ using Equation 8.

g. Update posterior probabilities $\mathbf{z}_{gk}$ using Equation A1.

h. Evaluate convergence by inspecting the difference between the log-likelihood at the current iteration $\text{LogL}_{\boldsymbol{\eta}_i}$ and the previous iteration $\text{LogL}_{\boldsymbol{\eta}_{i-1}}$. If convergence is not reached, repeat Steps 2a to 2h.

3. Repeat Steps 1 and 2 for a new random start until the required number of random starts is reached. From the converged solutions of the different random starts, select the solution with the highest $\text{LogL}_{\boldsymbol{\eta}}$ as the final solution.

Bias Correction

If the residual covariances for the endogenous factors are required to be *group-and-cluster-specific* (i.e., $\mathbf{\Psi}_{gk}^{\text{endog1}}$, $\mathbf{\Psi}_{gk}^{\text{endog2}}$), a correction in the algorithm is applied in Steps 2c to 2g to avoid bias in the regression parameters. Initially, we estimate cluster-specific residual covariances (Step 2d) and subsequently reconstruct the group-specific parts (Step 2e). This approach introduces bias in the regression parameters in $\mathbf{B}_k$ of the next iteration, as they are estimated without accounting for residual factor covariances that are group-and-cluster-specific (rather than cluster-specific).

To prevent this bias, one must change the weight w_{gk} used in Step 2c. Note that this correction starts from iteration two onward as it requires an initial estimation of the residual factor covariances $\mathbf{\Psi}_{gk}$. A step-by-step description of the correction is provided below, where at the beginning of each step, we indicate where it would be included in the procedure described in the Step 2:

1. In Step 2c, for each endogenous factor y, compute one weighted average factor covariance matrix $\mathbf{\Phi}_{ky}^{s1}$ per cluster k using Equation A3. Consider only the endogenous factor y and its predictors for each $\mathbf{\Phi}_{ky}^{s1}$. To determine the appropriate weights for each $\mathbf{\Phi}_{ky}^{s1}$, compute the weights w_{gky} per endogenous factor as $w_{gky} = \dfrac{\hat{\mathbf{z}}_{gky}N_g}{N}$. The posterior probabilities $\hat{\mathbf{z}}_{gky}$ are obtained from the original $\hat{\mathbf{z}}_{gk}$ as follows:
   a. In Step 2d, use the corresponding $\mathbf{\Phi}_{ky}^{s1}$ to estimate the SM for each endogenous factor y separately. Extract the resulting cluster-specific regression parameters $\mathbf{B}_{ky}$ from the separate SMs and assemble a complete $\mathbf{B}_k$ including all factors.
   b. In Step 2e, reconstruct the group-and-cluster-specific factor residual covariance matrix $\mathbf{\Psi}_{gk}$ using the same method as described in the Step 2. Let us denote the group-and-cluster-specific residual variance of each endogenous factor inside $\mathbf{\Psi}_{gk}$ as ψ_{gky}.
   c. In Step 2g, update the posterior probabilities $\hat{\mathbf{z}}_{gk}$ using Equation A1 as described in the Step 2. Per endogenous factor y, weigh the posteriors by the inverse of the endogenous factor's residual variance ψ_{gky}, that is, $\hat{\mathbf{z}}_{gky} = \dfrac{\hat{\mathbf{z}}_{gk}}{\psi_{gky}}$. In this way, create one $\mathbf{Z}_y$ matrix per endogenous factor. Note that the $\mathbf{Z}_y$ matrices are only used to correct the bias in the regression parameters but are not the final classification probabilities given to the user as part of the results.

Note that the bias correction described above is equivalent to a weighted least squares estimation of the regression parameters. In this case, we use the inverse of the factor residual variances ψ_{gky} to obtain adjusted posterior probabilities $\hat{\mathbf{z}}_{gky}$.

Appendix B

MI Testing Empirical Application

This appendix outlines the MI evaluation conducted on the data used for the empirical example in the Empirical Application section. The SEM model we used as a starting point can be seen in Figure 7. Before applying MMG-SEM, we needed to assess MI up to the metric invariance level (i.e., invariant loadings across groups), since this is necessary for comparing latent relations across groups. To this end, several MG-CFA's were conducted using the R package "lavaan" as described in Step 1 in the Method section. We deemed a configural invariant model acceptable if its fit index CFI was larger than 0.9 and RMSEA was lower than 0.08. Furthermore, a metric invariant model was acceptable if the drop in fit was not more than 0.02 for CFI and 0.01 for RMSEA (Rutkowski & Svetina, 2014). We used two measurement blocks (see the Empirical Application section): one for the emotions factors and one for the life satisfaction factor. Below, we report on the MI evaluation per measurement block.

MI of PE and NE

For the measurement block containing the two factors for the PE and NE, we first tested for configural invariance. Specifically, we ran an MG-CFA imposing, for each group, the MM for PE and NE as depicted in Figure 7, without equality restrictions on the measurement parameters. This resulted in an unacceptable model fit (CFI = 0.78, RMSEA = 0.11). Following the common recommendations in the literature (Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000), we inspected the modification indices

(*Appendices continue*)

Table B1

Step-by-Step Additions to the Measurement Model Based on Modification Indices

Parameter	CFI	RMSEA
Guilt ~~ shame	0.820	0.102
Worry ~~ stress	0.857	0.093
Pride ~~ gratitude	0.868	0.089
Gratitude ~~ love	0.880	0.086
PE =~ unpleasant	0.890	0.083
Pleasant ~~ unpleasant	0.897	0.081
NE =~ pride	0.903	0.079
Love ~~ jealousy	0.909	0.077
Pride ~~ anger	0.914	0.075
PE =~ sad	0.919	0.074

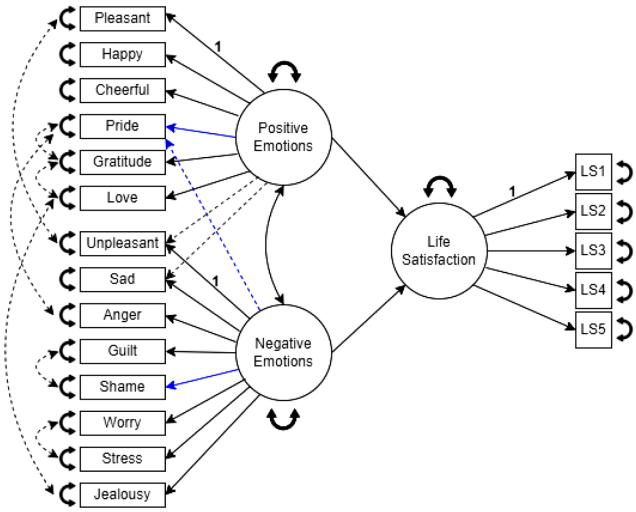
Note. ~~ represents a residual covariance between the items; =~ represents a factor loading; PE refers to the positive emotions factor; and NE refers to the negative emotions factor. CFI = comparative fit index; RMSEA = root mean square error of approximation.

to find parameters that would improve the fit of our model. We added 10 new measurement parameters (i.e., residual item covariances and cross-loadings) based on the modification indices. Note that they were not added all at once, but we included them one by one, each time re-evaluating the modification indices. In this way, we modified the model until the fit of the model was deemed acceptable, also considering whether modifications could be theoretically justified. Table B1 lists the new parameters added in the order of inclusion. After adding the mentioned parameters, the model fit was acceptable (CFI = 0.91, RMSEA = 0.07). Note that the modification indices also suggested adding the cross-loading NE =~ Gratitude. However, we decided to omit this modification index given that, after the model estimation, this cross-loading was moderately positive (i.e., $\lambda > 0.3$) for several countries, which was considered unjustifiable from a substantive point of view.

As our next step, we tested for metric invariance of the modified MM. After constraining the loadings to be equal across groups, the drop in fit exceeded the threshold of 0.02 for the CFI (i.e., CFI = 0.88, RMSEA = 0.078) and resulted in an unacceptable model fit. To examine partial metric invariance, we allowed certain factor loadings to be group-specific. To assess the improvement in model fit when freely estimating specific factor loadings, lavaan's `lavTestScore()` releases, one by one, each constrained loading for one group and performs a χ^2 test to check if there is a significant improvement in the model fit when freely estimating that loading for that group (Rosseel, 2012). Thus, it performs $J \times G$ χ^2 tests. We inspected for which loadings the independent χ^2 tests were significant most often (and, thus, for a large number of groups). Additionally, we checked whether the loading differences found when allowing these loadings to be group-specific corresponded to a nonignorable noninvariance size for at least one of the countries. The free factor loadings that would improve the model fit the most (indicated by the number of significant χ^2 tests in parentheses) were the following: NE =~ pride (7), PE =~ pride (7), and NE =~ shame (2). These loadings also account for nonignorable noninvariances since the maximal between-group difference (i.e., the difference between the smallest and largest loading) was 2.44, 1.45, and 1.50 for NE =~ pride, PE =~ pride, and NE =~ shame, respectively. The two loadings of pride were even positive in some groups

Figure B1

The SEM Model Used to Capture the Relations Between Emotions and Life Satisfaction After the Measurement Invariance Test Evaluation for the Two Measurement Blocks (i.e., Positive and Negative Emotions on the One Hand and Life Satisfaction on the Other Hand)



Note. SEM = structural equation modeling; LS = life satisfaction. See the online article for the color version of this figure.

and negative in others, which further underlines the importance of accounting for this noninvariance in the model. The resulting partially metric invariant model thus obtained an acceptable model fit that is very close to the fit of the configural invariant model (CFI = 0.90, RMSEA = 0.076).

MI of Life Satisfaction

For the life satisfaction factor, the configural invariant model had a good fit (CFI = 0.97, RMSEA = 0.08) and, thus, no modifications were needed. We proceeded by imposing equal loadings across groups to evaluate metric invariance. The metric invariant model presented an acceptable drop in terms of fit (CFI = 0.95, RMSEA = 0.08). Therefore, no factor loadings were freed to be group-specific, and the fully metric invariant model was accepted as the final model.

Final Model

The final model after all the modifications (i.e., combining the results of the two blocks) is depicted in Figure B1. The added measurement parameters can be seen in dashed lines, and the group-specific loadings for the emotions factors are presented in blue (dark gray). Lastly, the marker variables are identified by a "1" on top of the arrow of the corresponding factor loading.

Received July 7, 2023
Revision received February 28, 2024
Accepted March 25, 2024 ■