STRUCTURAL EQUATION MODELING
2025, VOL. 32, NO. 5
ISSN: 1070-5511 | E-ISSN: 1532-8007

OPEN ACCESS

On the Meaning of Measurement Invariance in Social Relations—Confirmatory Factor Analysis for Relative Variance Parameters

David Jendryczko and Fridtjof W. Nussbeck

University of Konstanz

ABSTRACT

We present and illustrate meaningful ways to assess relative variance parameters (variance components) in multiple indicator social relations—confirmatory factor analysis models for dyadic round-robin data where different types of measurement invariance may hold. With simulation studies, we investigate under which conditions of sample-size, true parameter values, and (mis-)specified invariance restrictions estimation issues as well as biased and inaccurate parameter estimates occur. Estimation issues are commonly observed in realistic data situations with low person-level variances and comparably few members per round-robin group. However, such issues can be effectively avoided by (falsely) implementing invariance restrictions across factor loadings without severely biasing relative variances for the sum-score and reciprocity correlations. Implications and limitations are discussed.

KEYWORDS

Dyadic data; measurement invariance; relative variances; round-robin; social relations

1. Introduction

The dyad (group of two) is the smallest unit in which humans—or any kind of animal for that matter—engage in social perception and behavior. Given that any person will be one half of many different dyads throughout life, the questions emerge: Why do different people perceive/behave differently in different dyads and how can we explain the dynamics within a dyad? With regard to perceiving others, specifically, one might ask: is it mainly interindividual differences in perceivers that determine how someone is seen? Or, is it mainly interindividual differences in how someone presents themselves in front of perceivers? Or, finally, will mostly the specific combination of the perceiver and the one to be perceived determine the perception?

In order to answer these and similar questions about the dynamics of dyads, researchers, since the late 1970s, have used the social relations model (SRM; Kenny, 2020; Kenny et al., 2006; Warner et al., 1979) as a conceptual and statistical tool. The SRM decomposes the variability in dyadic dynamics into actor- (e.g., perceiver), partner- (e.g., the one to be perceived or "target") and relationship- (e.g., the perceiver-target interaction) specific components and is increasingly used in almost all disciplines of the social and behavioral sciences (a frequently updated overview of the SRM-literature can be retrieved from David A. Kenny's website at: https://davidakenny.net/doc/srmbiblio.pdf). Yet, it is only due to rather recent methodological advancements (Jendryczko & Nussbeck, 2024; Nestler et al., 2020) that researchers can easily combine the SRM with flexible measurement-models within the structural equation modeling (SEM)-framework in order to separate measurement error from true dyadic dynamics and evaluate measurement-instruments specifically designed for assessing dyadic interactions. However, when such expanded Social Relations—Confirmatory Factor Analysis models (SR-CFAs) are used to investigate the sources of variability in social behavior and perceptions (actor-, partner-, and relationship-effects), the comparison of the sources heavily depends on the parameterization of the measurement.

The general aim of the current contribution is to discuss the impact of different parameterizations which is akin to the analysis of measurement-invariance. We will present psychometric tools and insights to properly assess this problem within SR-CFAs. To this end, we first explain the standard SRM and its extension to SR-CFA. Then, we discuss and illustrate the measurement-invariance problem and gain deeper understanding via a simulation study.

2. The Original Social Relations Model

The social relations model (SRM; Kenny, 2020; Kenny et al., 2006; Warner et al., 1979) is the standard tool for analyzing dyadic round-robin data. In the simplest case, round-robin data consist of one or several (round-robin) group(s) in which every member rates and is rated by every other member on one single indicator. Within every dyad of a group, every rating Y_{ij} of group-member i rating group-member j (e.g., how much does i like j?) and every opposite rating Y_{ji}

are decomposed in the following way:

$$\begin{aligned} Y_{ij} &= \kappa + \alpha_i + \beta_j + \gamma_{ij} \\ Y_{ji} &= \kappa + \alpha_j + \beta_i + \gamma_{ji}. \end{aligned} \tag{1}$$

Here κ denotes the constant intercept (overall mean for the rating). α_i and α_j denote the actor-effect (or rater-effect) for member i and j, respectively (denotes the degree to which member i/j tends to like other people in general). β_i and β_j denote the partner-effect (or target-effect) for member i and j, respectively (denotes how much member i/j tends to be generally liked by others). γ_{ij}/γ_{ji} denotes the relationship-effect or interaction-effect (denotes i's unique liking of j/j's unique liking of i). Importantly, the interaction-effect cannot be separated from measurement error. The latent variance-covariance structure (under the assumption of a multivariate normal distribution) is given by:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim MVN(0, \Sigma_{person}) \text{ with } \Sigma_{person} = \begin{pmatrix} \sigma^2_\alpha & \sigma_{\alpha\beta} \\ & \sigma^2_\beta \end{pmatrix} \text{ and}$$
$$\begin{pmatrix} \gamma_{ij} \\ \gamma_{ji} \end{pmatrix} \sim MVN(0, \Sigma_{dyad}) \text{ with } \Sigma_{dyad} = \begin{pmatrix} \sigma^2_\gamma & \sigma_{\gamma\gamma} \\ & \sigma^2_\gamma \end{pmatrix}. \tag{2}$$

Actor-effects and partner-effects vary across members denoted by the actor-effect variance σ^2_α and the partner-effect variance σ^2_β, respectively. The actor-partner covariance $\sigma_{\alpha\beta}$ denotes generalized reciprocity; for example, the degree to which "likers" are also generally liked by others. Relationship-effects vary across relationships denoted by the relationship variance σ^2_γ. The relationship covariance $\sigma_{\gamma\gamma}$ denotes dyadic reciprocity; for example, the degree to which a unique liking between i and j is mutual.

In many cases, researchers are particularly interested in relative variances (variance components) rather than in absolute variances as these depict the degree to which the observations depend on the different latent variables reflecting assimilation (A), consensus (C), and uniqueness (U), respectively:

$$\begin{aligned} A &= \sigma^2_\alpha/(\sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\gamma) \\ C &= \sigma^2_\beta/(\sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\gamma) \\ U &= \sigma^2_\gamma/(\sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\gamma). \end{aligned} \tag{3}$$

Assimilation denotes the degree to which the variability in dyadic interactions is attributable to the actor (e.g., the degree to which someone likes another person depends on the actor being a "liker"). Consensus denotes the degree to which variability in dyadic interactions is attributable to the partner (e.g., the degree to which someone likes another person depends on the partner being likeable). Uniqueness denotes the degree to which the variability in dyadic interactions is attributable to the unique relationships of actors and partners (e.g., the degree to which someone likes another person depends on the specific combination of the two persons). Each of these relative variances can only vary between 0 and 1 and they sum up to 1. Hence, the three relative variances can directly be interpreted as proportions or percentages if multiplied by 100. As noted before, the relationship effect cannot be separated from measurement error and will lead to overestimations of uniqueness components in many cases. In order to disentangle measurement error from relationship effects multiple indicators have to be administered and the simple ANOVA-based measurement decomposition has to be replaced relying on models of social relations—confirmatory factor analysis (SR-CFA), for example (Jendryczko & Nussbeck, 2024; Nestler et al., 2020). However, as these SR-CFA models introduce measurement models for all three latent variables (actor-effect, partner-effect, and relationship-effect) they may be specified with varying degrees of measurement invariance across the latent variables which in turn impacts how the relative variance parameters have to be computed in meaningful ways. We note that we use the term "measurement invariance" not in its strict traditional sense which implies the (in part) unchanged functionality of a measurement instrument across different populations or time-points (Little, 2013; Meredith, 1993). Rather, we use it to refer to restrictions on measurement-model parameters across different latent variables capturing different aspects of a general overarching construct (e.g., "being a liker", "being likable", "and liking/being liked by a specific person" as actor-, partner- and relationship-effect variables capturing different aspects of a general "liking"-construct).

Our objective is to (1) present and illustrate these meaningful ways of relative variance computation for different types of measurement-invariant SR-CFAs and (2) use simulation studies to investigate the impact of (falsely) implementing different types of measurement invariance on model-estimability, correct model detection, and parameter estimation bias and inaccuracy with a focus on relative variance parameters. As we will point out, such an investigation is necessary, as the implementation of (too) restrictive types of measurement invariance is needed for interpretative clarity and may be unavoidable for yielding admissible solutions in many cases.

In the remainder of this contribution, we will first present the SR-CFA model and the different types of measurement invariance that can be implemented. We will then point to necessary modifications of Equation (3) in order to compute relative variances for every type of measurement invariance. Models containing different types of measurement invariance are then illustrated with an empirical example. We will then point to issues of model estimation and report the results of a simulation study investigating the effects of model restrictions with an emphasis on relative variance components.

3. Social Relations—Confirmatory Factor Analysis

Imagine we assessed the construct of liking with three indicators: "I like this person." ($l = 1$), "I would like to get to know this person." ($l = 2$), and "I would like to become friends with this person." ($l = 3$), each measured on a scale from one to six (see Salazar Kämpf et al., 2018). Figure 1

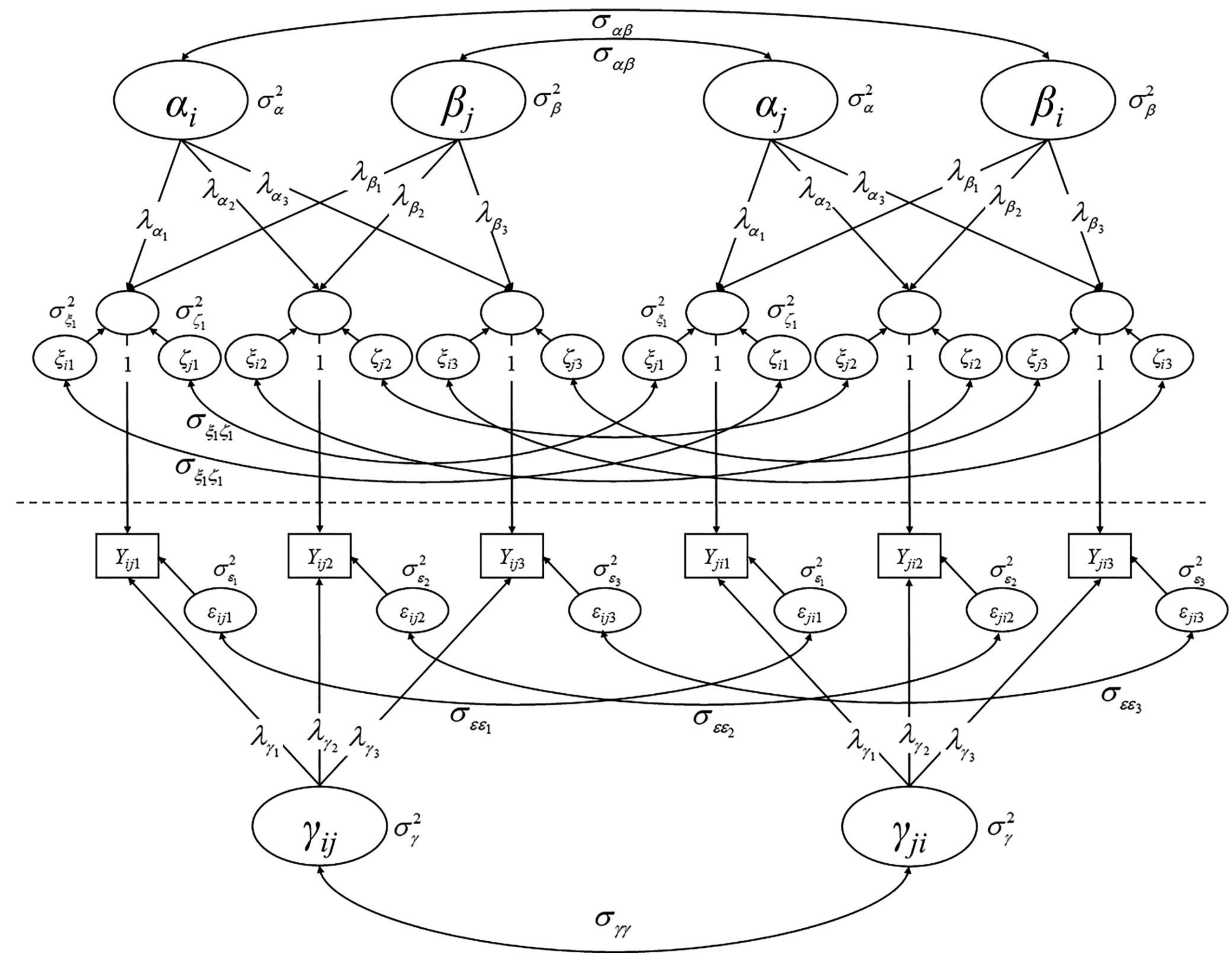

Figure 1. A configural invariant social relations—confirmatory factor analysis (SR-CFA) with three indicators. λ_{α_l} = factor loading of indicator *l* on latent actor-effect variable, λ_{β_l} = factor loading of indicator *l* on latent partner-effect variable, λ_{γ_l} = factor loading of indicator *l* on latent relationship-effect variable, σ^2_α = actor-effect variance, σ^2_β = partner-effect variance, σ^2_γ = relationship-effect variance, $\sigma_{\alpha\beta}$ = actor-partner covariance, $\sigma_{\gamma\gamma}$ = relationship-effect covariance, $\sigma^2_{\xi_l}$ = random actor-intercept variance for indicator *l*, $\sigma^2_{\zeta_l}$ = random partner-intercept variance for indicator *l*, $\sigma_{\xi_l\zeta_l}$ = random intercept covariance for indicator *l*, $\sigma^2_{\varepsilon_l}$ = residual relationship variance for indicator *l*, $\sigma_{\varepsilon\varepsilon_l}$ = residual relationship covariance for indicator *l*. Intercepts are not depicted. $\sigma^2_{\xi_l}$, $\sigma^2_{\zeta_l}$, and $\sigma_{\xi_l\zeta_l}$ are only labeled for the first indicator to maintain visual clarity. The person-level is depicted above the dashed horizontal line, the relationship-level is depicted below this line. Empty ovals indicate person-level aggregates (means) of round-robin indicators which are assessed at the relationship-level.

presents the SR-CFA model and the variance-covariance structure. The model equation expands to:

$$\begin{aligned} Y_{ijl} &= \kappa_l + \lambda_{\alpha_l}\alpha_i + \xi_{il} + \lambda_{\beta_l}\beta_j + \zeta_{jl} + \lambda_{\gamma_l}\gamma_{ij} + \varepsilon_{ijl} \\ Y_{jil} &= \kappa_l + \lambda_{\alpha_l}\alpha_j + \xi_{jl} + \lambda_{\beta_l}\beta_i + \zeta_{il} + \lambda_{\gamma_l}\gamma_{ji} + \varepsilon_{jil} \end{aligned} \tag{4}$$

(see Jendryczko & Nussbeck, 2024; Nestler et al., 2020). κ_l denotes the mean of indicator *l*. The meaning of α_i, β_i, and γ_{ij}[1] and their respective (co)variances remain unchanged from Equations (1) and (2), with the exception that they now represent the common variability across all indicators and, thus, the actor-, partner-, and relationship-effect of the construct of interest (general liking). That is, these latent variables reflect general tendencies to provoke low or high scores on the observed variables. λ_{α_l}, λ_{β_l}, and λ_{γ_l} depict factor loadings linking the respective latent variables to the indicators. ξ_{il} and ζ_{il} depict indicator-specific random actor- and partner-intercepts on person-level, respectively. That is, a person i might generally like others and be liked by others (high values in α_i and β_i, respectively), but—compared to others with the same scores on the latent variables—is more hesitant to endorse the item "I would like to become friend with that person" leading to a negative value in ξ_{i3} and may also evoke lower ratings on that item by others leading to a negative value in ζ_{i3}. For every person, these indicator-specific random intercepts may covary (denoted by $\sigma_{\xi_{il}\zeta_{il}}$). ε_{ijl} depicts the indicator-specific residual on relationship-level that cannot be explained by the latent variables. These indicator residuals may covary for every dyad (denoted by $\sigma_{\varepsilon\varepsilon_l}$). Note that ε_{ijl} also encapsulates measurement-error.

The model is configural invariant, meaning that a single construct (in our example "liking") is assumed to consist of two latent variables at the person-level (actor- and partner-effects) and one additional latent variable at the relationship-level (relationship-effect). Yet, the configural invariant model

---

[1] For the remainder, we only refer to one member (*i*) and one directional rating (*ij*) but all explanations, of course, also hold true when members switch their roles as actor and partner.

allows for different loading parameters on these three latent variables ($\lambda_{\alpha_l}, \lambda_{\beta_l}$ and λ_{γ_l} may differ). As a consequence, the variation in the latent variables affects indicators differently as it is weighed with the square of the loading parameter (see also Equation (8)). To make this clearer, assume that individuals A and B differ by one unit on their actor- and partner-effects, respectively. Then, the expected difference between A and B in the measurement by indicator l explained by the actor-effect would be λ_{α_l} and the expected difference between A and B in the measurement by indicator l explained by the partner-effect would be λ_{β_l}. If these loadings are not the same, the latent variables are measured on different scales and Equation (3) cannot be employed to derive meaningful relative variance parameters. Lastly, different random intercepts on person-level ξ_{il} and ζ_{il} mean that every person has a different "standard" for each specific expression of liking (different indicator).

In a weak invariant model, the factor loading for a specific indicator l is assumed to be equal for all three latent variables ($\lambda_{\alpha_l} = \lambda_{\beta_l} = \lambda_{\gamma_l} = \lambda_l$):

$$Y_{ijl} = \kappa_l + \lambda_l \alpha_i + \xi_{il} + \lambda_l \beta_j + \zeta_{jl} + \lambda_l \gamma_{ij} + \varepsilon_{ijl}. \tag{5}$$

This means that differences between actors, partners and relationships on the respective latent variables manifest to the same degree in the observations and Equation (3) can be applied. In a strong invariant model, indicator-specific random actor- and partner-intercepts are additionally removed implying that actor- and partner-effects, respectively, affect the manifest variables in the same way across individuals:

$$Y_{ijl} = \kappa_l + \lambda_l \alpha_i + \lambda_l \beta_j + \lambda_l \gamma_{ij} + \varepsilon_{ijl}. \tag{6}$$

The strong measurement invariant model bears the additional advantage that the latent variables are interpretable as conditional expectations in the framework of stochastic measurement theory (see Jendryczko & Nussbeck, 2024b, see also Koch et al., 2016).

4. Obtaining Meaningful Relative Variance-Parameters within SR-CFA

As already outlined above, we cannot compute a meaningful relative variance parameter for the latent variables α_i, β_j, and γ_{ij} within the configural invariant model due to these latent variables differing in their scales. However, we may compute the relative variances for individual indicators and their sum-score as these take differences in scaling into account by respecting the different factor-loadings. Within the configural invariant model, the variance for the round-robin observations of an indicator l is given by

$$Var(Y_{ijl}) = \lambda_{\alpha_l}^2 \sigma_\alpha^2 + \sigma_{\xi_l}^2 + \lambda_{\beta_l}^2 \sigma_\beta^2 + \sigma_{\zeta_l}^2 + \lambda_{\gamma_l}^2 \sigma_\gamma^2 + \sigma_{\varepsilon_l}^2 \tag{7}$$

because, in every directed rating (relationship), all latent variables are uncorrelated (see Figure 1). For constructing meaningful relative variance parameter estimates, we can focus on the part of the total variance that is determined by the three common factors ($\lambda_{\alpha_l}^2 \sigma_\alpha^2 + \lambda_{\beta_l}^2 \sigma_\beta^2 + \lambda_{\gamma_l}^2 \sigma_\gamma^2$) and define assimilation, consensus and uniqueness for every single indicator accordingly:

$$\begin{aligned} A_l &= \frac{\lambda_{\alpha_l}^2 \sigma_\alpha^2}{\lambda_{\alpha_l}^2 \sigma_\alpha^2 + \lambda_{\beta_l}^2 \sigma_\beta^2 + \lambda_{\gamma_l}^2 \sigma_\gamma^2} \\ C_l &= \frac{\lambda_{\beta_l}^2 \sigma_\beta^2}{\lambda_{\alpha_l}^2 \sigma_\alpha^2 + \lambda_{\beta_l}^2 \sigma_\beta^2 + \lambda_{\gamma_l}^2 \sigma_\gamma^2} \\ U_l &= \frac{\lambda_{\gamma_l}^2 \sigma_\gamma^2}{\lambda_{\alpha_l}^2 \sigma_\alpha^2 + \lambda_{\beta_l}^2 \sigma_\beta^2 + \lambda_{\gamma_l}^2 \sigma_\gamma^2}. \end{aligned} \tag{8}$$

If either weak or strong invariance holds, these equations simplify to

$$\begin{aligned} A_l &= \frac{\lambda_l^2 \sigma_\alpha^2}{\lambda_l^2 \sigma_\alpha^2 + \lambda_l^2 \sigma_\beta^2 + \lambda_l^2 \sigma_\gamma^2} = \frac{\lambda_l^2 \sigma_\alpha^2}{\lambda_l^2 (\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2)} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2} \\ C_l &= \frac{\lambda_l^2 \sigma_\beta^2}{\lambda_l^2 \sigma_\alpha^2 + \lambda_l^2 \sigma_\beta^2 + \lambda_l^2 \sigma_\gamma^2} = \frac{\lambda_l^2 \sigma_\beta^2}{\lambda_l^2 (\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2)} = \frac{\sigma_\beta^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2} \\ U_l &= \frac{\lambda_l^2 \sigma_\gamma^2}{\lambda_l^2 \sigma_\alpha^2 + \lambda_l^2 \sigma_\beta^2 + \lambda_l^2 \sigma_\gamma^2} = \frac{\lambda_l^2 \sigma_\gamma^2}{\lambda_l^2 (\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2)} = \frac{\sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2}, \end{aligned} \tag{9}$$

automatically implying that all indicators have the same degree of assimilation, consensus, and uniqueness and making Equation (3) applicable again since latent variables are now measured on the same scale.

We may also aggregate across all indicators and determine relative variance parameters for a sum-score $Y_{ij\bullet}$ which has the following model-implied sum:

$$Y_{ij\bullet} = \sum_{1 \le l \le p} Y_{ijl} = \sum_{1 \le l \le p} \mu_l + \lambda_{\alpha_l} \alpha_i + \xi_{il} + \lambda_{\beta_l} \beta_j + \zeta_{jl} + \lambda_{\gamma_l} \gamma_{ij} + \varepsilon_{ijl}. \tag{10}$$

Here, p denotes the total number of indicators. For the configural invariant model, the variance of a sum-score is given by:

$$\begin{aligned} Var(Y_{ij\bullet}) = &\sum_{1 \le l \le p} \lambda_{\alpha_l}^2 \sigma_\alpha^2 + \lambda_{\beta_l}^2 \sigma_\beta^2 + \lambda_{\gamma_l}^2 \sigma_\gamma^2 + \sigma_{\xi_l}^2 + \sigma_{\zeta_l}^2 + \sigma_{\varepsilon_l}^2 \\ &+ \sum_{1 \le q \le p-1} \sum_{q+1 \le r \le p} 2\lambda_{\alpha_q} \lambda_{\alpha_r} \sigma_\alpha^2 + 2\lambda_{\beta_q} \lambda_{\beta_r} \sigma_\beta^2 \\ &+ 2\lambda_{\gamma_q} \lambda_{\gamma_r} \sigma_\gamma^2. \end{aligned} \tag{11}$$

The summands in Equation (11) that contain a multiplicative of 2 respect model-implied covariances between different indicators. These covariances only emerge due to loading on the same latent variable (see Figure 1). Assimilation for the sum-score is thus given by:

$$\begin{aligned} A_\bullet &= \frac{\sum_{1 \le l \le p} \lambda_{\alpha_l}^2 \sigma_\alpha^2 + \sum_{1 \le q \le p-1} \sum_{q+1 \le r \le p} 2\lambda_{\alpha_q} \lambda_{\alpha_r} \sigma_\alpha^2}{\sum_{1 \le l \le p} \lambda_{\alpha_l}^2 \sigma_\alpha^2 + \lambda_{\beta_l}^2 \sigma_\beta^2 + \lambda_{\gamma_l}^2 \sigma_\gamma^2 + \sum_{1 \le q \le p-1} \sum_{q+1 \le r \le p} 2\lambda_{\alpha_q} \lambda_{\alpha_r} \sigma_\alpha^2 + 2\lambda_{\beta_q} \lambda_{\beta_r} \sigma_\beta^2 + 2\lambda_{\gamma_q} \lambda_{\gamma_r} \sigma_\gamma^2} \\ &= \frac{\sigma_\alpha^2 (\sum_{1 \le l \le p} \lambda_{\alpha_l}^2 + \sum_{1 \le q \le p-1} \sum_{q+1 \le r \le p} 2\lambda_{\alpha_q} \lambda_{\alpha_r})}{\sigma_\alpha^2 (\sum_{1 \le l \le p} \lambda_{\alpha_l}^2 + \sum_{1 \le q \le p-1} \sum_{q+1 \le r \le p} 2\lambda_{\alpha_q} \lambda_{\alpha_r}) + \sigma_\beta^2 (\sum_{1 \le l \le p} \lambda_{\beta_l}^2 + \sum_{1 \le q \le p-1} \sum_{q+1 \le r \le p} 2\lambda_{\beta_q} \lambda_{\beta_r}) + \sigma_\gamma^2 (\sum_{1 \le l \le p} \lambda_{\gamma_l}^2 + \sum_{1 \le q \le p-1} \sum_{q+1 \le r \le p} 2\lambda_{\gamma_q} \lambda_{\gamma_r})}. \end{aligned} \tag{12}$$

For consensus ($C_\bullet$) and uniqueness ($U_\bullet$) of the sum-score, each α in the enumerator of Equation (12) must be replaced by β or γ, respectively. For the weak or strong invariant model, this simplifies to

$$\begin{aligned} A_\bullet &= \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2} \\ C_\bullet &= \frac{\sigma_\beta^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2} \\ U_\bullet &= \frac{\sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2}, \end{aligned} \tag{13}$$

implying that the relative variances for the sum score are identical to the relative variances for each indicator. Hence, Equation (12) can be simplified yielding Equation (3)/ Equation (13). In Appendix A, we exemplify the derivation of Equation (13) from Equation (12) for two round-robin indicators.

5. Empirical Illustration

We illustrate the interpretation of relative variance parameters in SR-CFAs under different assumptions about measurement invariance using open-access round-robin data from a study by Salazar Kämpf (2018; available at https://osf.io/asu4g/). In particular, we estimated (using full information maximum likelihood) a configural, a weak, and a strong invariant model for the round-robin data of mimicry and compared them via standard likelihood-ratio (i.e., χ^2-difference) tests for SEMs.

Each of a total of 139 participants was randomly assigned to one of 26 round-robin groups containing either four (4 groups), five (9 groups), or six (13 groups) members. Each dyad of each group was video-recorded during a conversation. Three independent raters (academic staff) watched these videos and rated how much each person imitated the other person during the conversation on a scale from 1 to 6 (higher values correspond to more mimicry). Thus, each of the three independent raters ($l = 1$, 2, or 3) can be understood as an indicator for actor i's mimicry of partner j: Y_{ijl} (see Equation (4)). For more information, we refer to Salazar Kämpf et al. (2018).

The data-analysis was executed in R (R core team, 2018) with the OpenMx-package (Neale et al., 2016) using the wide-data approach as explained in Jendryczko & Nussbeck (2024). The supplementary material (https://osf.io/rje2n/) contains the files required for a replication.

5.1. Results

Table 1 displays the results for absolute model parameters and relative variance parameters across the three models. The configural and the weak invariant model contained anomalous results. In the configural invariant model, the random-intercept partner-variance for the second indicator was estimated to be negative ($\sigma_{\zeta_2}^2 = -0.021$). The weak invariant model had the same problem ($\sigma_{\zeta_2}^2 = -0.003$) and also yielded a random-intercept actor-partner correlation outside the parameter-space for the first indicator ($\rho_{\xi_1\zeta_1} = -2.645$). It should be noted that the negative variances were not significantly different from zero and that the results of both models are quite comparable to the results of the strong invariant model which did not yield any anomalous results. Yet, researches should interpret models with anomalous results cautiously if at all.

In the configural invariant model, the factor loadings were quite different between the latent actor-, partner-, and relationship-effect—especially for the second indicator ($\lambda_{\alpha_2} = 0.758$, $p < 0.001$; $\lambda_{\beta_2} = 1.173$, $p < 0.001$; $\lambda_{\gamma_2} = 0.698$, $p < 0.001$). This shows that the latent variables were not measured on the same scale. Two participants with a difference of 1 in the latent tendency to imitate (α) are predicted to elicit a difference of 0.758 scale-values in the judgement of the second rater. Yet, if two participants have the same difference of 1 in latent imitability (β), they would be expected to differ by even 1.173 scale-values in the judgement of the second rater. Moreover, if two participants differ by 1 in latent imitativeness of a specific person or latent imitability by a specific person (γ), they would be expected to differ by only 0.698 scale-values in the judgement of the second rater. It follows that the variances of the latent actor-, partner-, and relationship-effects cannot be compared directly and must be weighed by these different factor loadings. Since the differences between factor-loadings were also quite different across the three latent variables (e.g., $\lambda_{\alpha_2} - \lambda_{\alpha_3} = 0.758 - 0.843 = -0.085$; $\lambda_{\beta_2} - \lambda_{\beta_3} = 1.173 - 0.761 = 0.412$; $\lambda_{\gamma_2} - \lambda_{\gamma_3} = 0.698 - 0.720 = -0.022$), different relative variance parameters were yielded for the different indicators. Yet, they all have in common that the most variability in mimicry can be explained by the specific combination of actor and partner, followed by characteristics of the actor with only little variability accounted for by characteristics of the partner. This is also reflected in sum-score uniqueness ($U_\bullet = 0.508$), assimilation ($A_\bullet = 0.410$), and consensus ($C_\bullet = 0.082$). Thus, if we assume only configural invariance to hold and take the sum-score across the judgements of the three sample-independent raters to represent a quantitative display of imitative behavior, we yield the following conclusions: half of the variability (50.8%) in the mimicry ratings is due to the particular dyad (uniqueness), 41.0% can be explained by the mimicking (imitating) actor (assimilation) and only 8.2% are attributable at the partner evoking imitative behavior (consensus).

Even though factor-loadings were substantially different in the configural invariant model, the likelihood-ratio test comparing the model with the weak invariant model suggested that the assumption of equal factor loadings does not have to be rejected (χ^2 (4) $= 6.164$, $p = 0.187$). Due to the small sample size and, therefore, quite large standard errors of the factor-loadings, we are well-advised to interpret this result with caution. Under the assumption of weak invariance (i.e., the second and third factor-loading are, respectively, equal across all latent variables: $\lambda_2 = 0.759$, $p < 0.001$; $\lambda_3 = 0.752$, $p < 0.001$), we can observe that the rank-order of latent variables with regard to relative amount of total

Table 1. Results of the configural, weak, and strong invariant model from the empirical illustration.

	Model								
	Configural			Weak			Strong		
Indicator	1	2	3	1	2	3	1	2	3
Intercept									
κ_I	2.976***	2.902***	3.033***	2.976***	2.902***	3.033***	2.975***	2.903***	3.031***
Loading									
λ_{α_I}	1	0.758***	0.843***	1	0.759***	0.752***	1	0.772***	0.786***
λ_{β_I}	1	1.173***	0.761***	1	0.759***	0.752***	1	0.772***	0.786***
λ_{γ_I}	1	0.698***	0.720***	1	0.759***	0.752***	1	0.772***	0.786***
Variance									
σ^2_α		0.516***			0.523***			0.569***	
σ^2_β		0.081			0.135**			0.138**	
σ^2_γ		0.738***			0.659***			0.592***	
A_I	0.387	0.387	0.461	0.397	0.397	0.397	0.438	0.438	0.438
C_I	0.060	0.145	0.059	0.102	0.102	0.102	0.106	0.106	0.106
U_I	0.553	0.469	0.481	0.500	0.500	0.500	0.455	0.455	0.455
$A_\bullet$		0.410			0.397			0.438	
$C_\bullet$		0.082			0.102			0.106	
$U_\bullet$		0.508			0.500			0.455	
$\sigma^2_{\xi_I}$	0.040	0.089**	0.175***	0.027	0.092***	0.200***			
$\sigma^2_{\zeta_I}$	0.011	**−0.021**	0.015	0.006	**−0.003**	0.015			
$\sigma^2_{\varepsilon_I}$	0.346***	0.264***	0.320***	0.368***	0.251***	0.319***	0.433***	0.335***	0.500***
Covariance									
$\sigma_{\alpha\beta}$		0.156** (0.766)			0.201*** (0.757)			0.197*** (0.702)	
$\sigma_{\gamma\gamma}$		0.554*** (0.750)			0.492*** (0.747)			0.479*** (0.809)	
$\sigma_{\xi_I\zeta_I}$	−0.019 (−0.903)	**0.017** **(*NA*)**	−0.023 (−0.443)	**−0.033** **(−2.645)**	**0.027*** **(*NA*)**	−0.025 (−0.461)			
$\sigma_{\varepsilon\varepsilon_I}$	0.095* (0.273)	0.017 (0.063)	0.049 (0.154)	0.104* (0.283)	0.009 (0.038)	0.049 (0.153)	0.072 (0.166)	0.036 (0.108)	0.005 (0.009)

Notes. Values of 1 were fixed for model identification. Numbers in parenthesis depict correlations. Values printed in bold reflect inadmissible solutions, i.e., either a negative variance, a correlation with an absolute value larger than 1, or a covariance/correlation between two variables of which at least one has a negative variance. α = general actor-effect variable, β = general partner-effect variable, γ = general relationship-effect variable, ξ_I = random-intercept actor-effect variable for indicator I, ζ_I = random-intercept partner-effect variable for indicator I, ε_I = residual for indicator I (encapsulates measurement-error), A = assimilation, C = consensus, U = uniqueness. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

variance explained remains unchanged, yet the proportions change quite a bit—depending on the indicator. For example, in comparison to the configural invariant model, assimilation for the third indicator decreases by 6.4 percentage points, consensus increases by 4.3 percentage points, and uniqueness increases by 1.9 percentage points in the weak invariant model. Moreover, the factor-loading for the common partner-effect on the second indicator underwent the strongest decrease from configural to weak invariance (from 1.173 to 0.759) and, accordingly, also underwent a strong decrease in consensus (from 14.5% to 10.2%). Changes in the variance-decomposition of the sum-score, however, were much less pronounced between the models ($A_{\bullet weak} - A_{\bullet configural} = 0.013$, $C_{\bullet weak} - C_{\bullet configural} = 0.020$, $A_{\bullet weak} - A_{\bullet configural} = -0.008$). If we assume weak invariance to hold, then the latent variables (α, β, and γ) represent quantitative differences in mimicry on the same scale and we, again, conclude that around two fifth (39.7%) of the variability in mimicry within dyadic dynamics is explainable by interindividual differences in imitativeness, while merely around one tenth (10.2%) of this variability is explainable by interindividual differences in imitability. Half of the variability (50.0%) cannot be accounted for by interindividual differences and is attributed to differences in dyadic interactions.

It is also noteworthy that the reciprocity correlations only change slightly with the imposition of weak measurement invariance (configural invariant model: $\rho_{\alpha\beta} = 0.766$, $\rho_{\gamma\gamma} = 0.750$; weak invariant model: $\rho_{\alpha\beta} = 0.757$, $\rho_{\gamma\gamma} = 0.747$; all $p < 0.010$). Thus, whether we assume configural or weak invariance, we conclude that highly imitable individuals strongly tend to be reciprocally imitative and that (beyond interindividual differences in imitativeness and imitability) imitative behavior within a dyad tends to be strongly reciprocal.

The likelihood-ratio test between the weak invariant and the strong invariant model revealed that strong invariance should be rejected (χ^2 (9) = 132.430, $p < 0.001$). In comparison to the other two models, we still observe the same rank-order of relative variances in the strong invariant model, yet assimilation and uniqueness are now much closer to each other and the changes in comparison to previous models are stronger ($A_\bullet = 0.438$, 4.1 percentage points increase in comparison to the weak invariant model; $C_\bullet = 0.106$, 0.4 percentage points increase in comparison to the weak invariant model; $U_\bullet = 0.455$, 4.5 percentage points decrease in comparison to the weak invariant model). Thus, under the assumption of strong measurement invariance, more variability in dyadic dynamics is explainable by interindividual differences in imitativeness, elevating the relative amount of variance explained by interindividual differences to over 50%. The reciprocity correlations are also quite different in comparison to the previous models ($\rho_{\alpha\beta} = 0.702$, $\rho_{\gamma\gamma} = 0.809$; all $p < 0.001$).

6. Statistical Benefits of Stricter Types of Measurement Invariance and the Concrete Aims of the Simulation Study

To the best of our knowledge, there exist, as of this writing, only two real-data applications of an SR-CFA in the literature (apart from the empirical illustration above) and relative variances and measurement invariance were not the focus in either one of them. The first mainly serves illustration purposes and is presented in the seminal paper by Nestler et al. (2020). The second (Jendryczko & Nussbeck, 2024b) is a re-analysis of the same open-access round-robin data-set used in the current contribution. It includes liking, meta-liking (the belief to be liked) and mimicry measures (Salazar Kämpf et al., 2018). We want to draw the attention of future applicants of SR-CFAs to statistical problems they are likely to encounter (see Jendryczko & Nussbeck, 2024b) and discuss potential solutions and implications. To this end, we point out that an SR-CFA (and the standard SRM) holds conceptual and statistical similarities to standard two-level (CFA) models, as variability in dyadic-round robin ratings is decomposed into a person-level (comprising variability of persons/members as actors and partners) and a relationship-level (comprising the variability of the relationships). Two-level CFAs (and, generally, multilevel CFAs) are notorious for yielding estimation issues such as inadmissible solutions in the shape of negative variance estimates (see also the empirical illustration above) and/or irregular loading patterns, especially when the number of sampled clusters (higher level-units) and/or variability on higher levels is small (Jak, 2019; Jak et al., 2014; Li & Beretvas, 2013; Lüdtke, et al., 2011). Such issues can often be resolved by implementing stricter forms of measurement invariance on the multilevel measurement structure (Depaoli & Clifton, 2015; González-Romá & Hernández, 2017; Kim & Cao, 2015).

On the basis of this, the current contribution presents simulation studies with the following purposes: (1) we investigate whether such estimation issues also occur in SR-CFA. Such an investigation is important for two reasons. First, round-robin data-inquiry is difficult to conduct and often bound to a rather small sample size of round-robin group members. Large numbers of group members can yield estimation issues by themselves (Jendryczko & Nussbeck, 2024a) and the work-load for each study-participant increases exponentially with every additional group-member that needs to be rated, especially when multiple indicators are used (or even multiple constructs with multiple indicators are assessed). Second, findings from the previous SRM-literature suggest that many psychological constructs assessed with round-robin ratings indeed display large relationship-variance but comparatively low actor- and/or partner-variance on the higher person-level (see Kenny, 2020 for an overview; see https://davidakenny.net/doc/srmbiblio.pdf for a collection of SRM references provided by David A. Kenny). (2) We investigate whether such estimation issues can be resolved by implementing various types of measurement invariance—which, at the same time, solve the issue of rendering actor-, partner-, and relationship-effects comparable (see above). (3) We investigate the degree to which implementing too restrictive types of invariance imposes estimation biases and inaccuracies in parameter estimates—with a pronounced focus on relative variance parameters. While it is to be expected that some absolute parameter estimates will suffer biases and inaccuracies in such scenarios, it is less clear how relative variance parameters will be affected. As these relative variances are often considered important from a substantive perspective, their robustness against invariance-violations should be investigated, especially if the violations are unavoidable for yielding admissible solutions and clearer interpretations of latent variables.

7. Design of the Simulation Study

We specified two conditions (two branches in the following) of true population parameters with respect to the relative variance components. For these two branches, a configural, a weak, and a strong invariant true population model was defined (see Table 2). For branch 1, population parameters were chosen so that they reflect an "ideal" case, in which all true parameter values are substantial and, thus, less expected to yield estimation issues. For branch 2, we aimed at reflecting a more "realistic" case, in which the actor- and partner-variance were substantially lower than the relationship variance, with the actor-effect variance being especially low. All true parameter values for branch 2 were inspired by the previously mentioned re-analysis (Jendryczko & Nussbeck, 2024b) of the open-access data set provided by Salazar Kämpf et al. (2018; available at https://osf.io/asu4g/). Importantly, the differences between factor loadings related to different latent variables were kept substantial in the configural invariant models of both branches, so that violations of configural invariance are sufficiently pronounced.

Each of the three true models were estimated with every model of invariance as defined previously. There were twelve conditions of sample size. The number of round-robin groups was either $M = 25$, $M = 50$, or $M = 100$, and the number of members per round-robin group was either $N = 5$, $N = 8$, $N = 10$, or $N = 15$.

Estimated models were compared via likelihood-ratio (LR) tests. We always compared the configural invariant model (most general model) with the weak invariant model (restricted model) and the weak invariant model (general model) with the strong invariant model (most restricted model), yielding 4 and 9 degrees of freedom, respectively. If the p-value associated to the test-statistic was equal to or below 0.05, the more general models was maintained; otherwise, the more restrictive model was accepted.

For the parameter estimates, we investigated the dependent variables of relative (parameter estimation) bias (RB), relative (parameter estimation) inaccuracy (RI) and coverage rate (CR). RB is given by

$$RB_\theta = \frac{\sum_{1 \leq g \leq h} \frac{\hat{\theta}_g - \theta}{\theta}}{h}, \tag{14}$$

with $\hat{\theta}_g$ denoting the estimate of a parameter in the current sample g, θ denoting the true parameter value and h

Table 2. Population parameters in the simulation study (two branches with different population parameters).

Parameter	Branch 1			Branch 2		
	Configural	Weak	Strong	Configural	Weak	Strong
$\lambda_{\alpha_1} = \lambda_{\beta_1} = \lambda_{\gamma_1}$	1	1	1	1	1	1
$\lambda_{\alpha_2} = \lambda_{\alpha_3}$	1	1	1	1	1	1
$\lambda_{\beta_2} = \lambda_{\beta_3}$	1.4	1	1	1.4	1	1
$\lambda_{\gamma_2} = \lambda_{\gamma_3}$	1.2	1	1	1.2	1	1
σ^2_α	0.8	0.8	0.8	0.041	0.041	0.041
σ^2_β	0.6	0.6	0.6	0.282	0.282	0.282
σ^2_γ	1	1	1	0.630	0.630	0.630
$\sigma_{\alpha\beta}$	0.208 (0.3)	0.208 (0.3)	0.208 (0.3)	0.002 (0.015)	0.002 (0.015)	0.002 (0.015)
$\sigma_{\gamma\gamma}$	0.5 (0.5)	0.5 (0.5)	0.5 (0.5)	0.169 (0.268)	0.169 (0.268)	0.169 (0.268)
$\sigma^2_{\xi_1}; \sigma^2_{\xi_2}; \sigma^2_{\xi_3}$	all 0.2	all 0.2	all 0	0.038; 0.022; 0.082	0.038; 0.022; 0.082	all 0
$\sigma^2_{\zeta_1}; \sigma^2_{\zeta_2}; \sigma^2_{\zeta_3}$	all 0.294	all 0.294	all 0	0.015; 0.017; 0.017	0.015; 0.017; 0.017	all 0
$\sigma_{\xi_1\zeta_1}; \sigma_{\xi_2\zeta_2}; \sigma_{\xi_3\zeta_3}$	all 0.024 (0.1)	all 0.024 (0.1)	all 0	0.010 (0.423); 0.003 (0.132); 0.003 (0.069)	0.010 (0.423); 0.003 (0.132); 0.003 (0.069)	all 0
$\sigma^2_{\varepsilon_1}; \sigma^2_{\varepsilon_2}; \sigma^2_{\varepsilon_3}$	all 0.36	all 0.36	all 0.36	0.255; 0.158; 0.303	0.255; 0.158; 0.303	0.255; 0.158; 0.303
$\sigma_{\varepsilon\varepsilon_1}; \sigma_{\varepsilon\varepsilon_2}; \sigma_{\varepsilon\varepsilon_3}$	all 0.036 (0.1)	all 0.036 (0.1)	all 0.036 (0.1)	−0.005 (−0.019); −0.009 (−0.058); −0.026 (−0.087)	−0.005 (−0.019); −0.009 (−0.058); −0.026 (−0.087)	−0.005 (−0.019); −0.009 (−0.058); −0.026 (−0.087)

Notes. λ_{α_l} = factor laoding of indicator *l* on latent actor-effect variable, λ_{β_l} = factor laoding of indicator *l* on latent partner-effect variable, λ_{γ_l} = factor laoding of indicator *l* on latent relationship-effect variable, σ^2_α = actor-effect variance, σ^2_β = partner-effect variance, σ^2_γ = relationship-effect variance, $\sigma_{\alpha\beta}$ = actor-partner covariance, $\sigma_{\gamma\gamma}$ = relationship-effect covariance, $\sigma^2_{\xi_l}$ = random actor-intercept variance for indicator *l*, $\sigma^2_{\zeta_l}$ = random partner-intercept variance for indicator *l*, $\sigma_{\xi_l\zeta_l}$ = random intercept covariance for indicator *l*, $\sigma^2_{\varepsilon_l}$ = residual relationship variance for indicator *l*, $\sigma_{\varepsilon\varepsilon_l}$ = residual relationship covariance for indicator *l*. Constant intercepts were set to zero. Numbers in parenthesis next to covariance parameters indicate correlations.

denoting the total number of samples. For the RI (Kerkhoff & Nussbeck, 2022), the absolute value of the difference between estimate and true value is taken:

$$RI_\theta = \frac{\sum_{1\le g\le h}\frac{|\hat{\theta}_g-\theta|}{\theta}}{h}. \tag{15}$$

The CR reflects the relative number of samples for which the 95%-confidence interval (CI) of a parameter estimate contains the true value. Absolute values for RBs should be below 0.10, CR values should be between 0.91 and 0.98, and RIs should reduce with increasing sample size (Muthén & Muthén, 2002).

For the relative variances and reciprocity correlations ($\rho_{\alpha\beta}$ and $\rho_{\gamma\gamma}$), absolute bias and absolute inaccuracy was computed; meaning that the difference between sample estimate and true value was not divided by the true value. Since relative variances can only range from zero to one and always sum up to one and correlations can only range from minus one to one, absolute deviations hold more convenient interpretations for these parameters.

For every condition with a specific sample size and true model, 1500 samples were drawn. Samples with inadmissible solutions were discarded for further analyses. Solutions were defined as in-admissible when standard errors could not be computed or were very large (> 1), factor loading patterns were irregular (e.g., some factor loadings were negative or larger than 3), or variance estimates were negative. Samples with multivariate normally distributed data were drawn using the mvtnorm package version 1.1.1 (Genz et al., 2020). Model estimations were executed with the OpenMx package version 2.19.5 (Neale et al., 2016) in R (R Core Team, 2018) using the full information maximum likelihood method with the wide-data approach as presented in Jendryczko & Nussbeck (2024). The supplementary material (https://osf.io/rje2n/) contains R-scripts for the setup of the simulation study and the results.

8. Results

8.1. Inadmissible Solutions

In branch 1, inadmissible solutions were rare and only found in two conditions of sample size. In the condition $N \times M = 5 \times 25$, 1.47% of samples yielded inadmissible solutions when the configural model was true and estimated, 0.07% when the configural model was true and the weak model was estimated, and 1.07% when the weak model was true and the configural model was estimated. With 5×50, 0.07% of samples yielded inadmissible solutions when the configural model was true and estimated.

Table 3 shows the proportions of inadmissible solutions for branch 2. All models are generally more estimable the more restrictive the true model is. With small sample sizes, the configural and weak invariant models were barely estimable. Increasing the number of members per round-robin group was more effective than increasing the number of round-robin groups. Inadmissible solutions can be kept under 10% for the estimation of all models (regardless of the true model) when the number of participants is either 8×100, 10×50, 10×100, 15×25, or 15×100. All following results are based on admissible solutions.

8.2. Likelihood Ratio Tests

Table 4 presents the rates for detecting (in)correctly specified models of invariance via likelihood ratio tests for both branches. Substantial issues only occurred when the true model was the configural invariant model and an LR-test comparing the configural and the weak invariant model was computed. Here, sufficiently large sample sizes are needed. With a sample size of 5×25, only 39.9% of tests in branch 1 and only 26.4% of tests in branch 2 correctly identified the weak invariant model as too restrictive. For a sample size of 5×50, these numbers were 71.7% and 46.2%, respectively and for a sample size of 8×25, these numbers

Table 3. Proportions of inadmissible solutions (relative to 1500) for various conditions of true model, estimated model and sample size in the second branch of the simulation study.

	True model								
	Configural			Weak			Strong		
	Estimated model								
$N \times M$	Configural	Weak	Strong	Configural	Weak	Strong	Configural	Weak	Strong
5×25	0.927	0.656	0.122	0.875	0.610	0.082	0.503	0.118	0.118
5×50	0.786	0.417	0.040	0.687	0.339	0.024	0.416	0.068	0.067
5×100	0.495	0.135	0.008	0.427	0.115	0.001	0.274	0.010	0.009
8×25	0.479	0.107	0.005	0.383	0.072	0.001	0.216	0.003	0.005
8×50	0.246	0.016	0	0.129	0.01	0	0.138	0	0
8×100	0.065	0	0	0.035	0	0	0.065	0	0
10×25	0.234	0.009	0	0.199	0.012	0	0.199	0	0
10×50	0.089	0.001	0	0.035	0	0	0.047	0	0
10×100	0.033	0	0	0.017	0	0	0.078	0	0
15×25	0.077	0	0	0.013	0.001	0	0.068	0	0
15×50	0.035	0	0	0.013	0	0	0.106	0	0
15×100	0.193	0	0	0.001	0	0	0.069	0	0

Notes. $N \times M$ = sample size comprising N group-members and M round-robin groups.

Table 4. Proportions of correctly detected (mis-)specifications of invariance via likelihood-ratio tests ($\alpha = 0.05$) in the simulation study.

	Branch 1				Branch 2			
	C vs. W		W vs. S		C vs. W		W vs. S	
$N \times M$	True: C	True: W	True: W	True: S	True: C	True: W	True: W	True: S
5×25	0.399	0.943	1	0.937	0.264	0.969	0.993	0.934
5×50	0.717	0.944	1	0.945	0.462	0.961	1	0.952
5×100	0.966	0.963	1	0.951	0.843	0.966	1	0.955
8×25	0.783	0.953	1	0.945	0.763	0.967	1	0.947
8×50	0.989	0.951	1	0.953	0.969	0.956	1	0.951
8×100	1	0.949	1	0.954	1	0.949	1	0.952
10×25	0.918	0.935	1	0.949	0.939	0.950	1	0.949
10×50	0.997	0.959	1	0.953	0.999	0.959	1	0.952
10×100	1	0.945	1	0.949	1	0.945	1	0.949
15×25	0.997	0.954	1	0.952	1	0.956	1	0.954
15×50	1	0.951	1	0.942	1	0.954	1	0.945
15×100	1	0.953	1	0.955	1	0.953	1	0.959

Notes. $N \times M$ = sample size comprising N group-members and M round-robin groups, C = configural model, W = weak model, S = strong model. A "vs." indicates a likelihood-ratio test of two models. The degrees of freedom were 4 for a C vs. W test and 9 for a W vs. S test.

were 78.3% and 76.3%, respectively. For all other conditions of sample size this detection rate was above 80% in both branches.

8.3. Absolute Parameters

While we focus on absolute bias and absolute inaccuracies for the relative variance parameters and reciprocity correlations, the supplementary material contains the complete results for absolute parameters in the form of digital tables as well as various large graphical presentations of RB, RI and CR for every model parameter in every condition. Table 5 displays some summary statistics of RB, RI and CR for cases in which the estimated model was equal to or more general than the true model for both branches. Overall, the parameter recovery worked quite well in those cases. In branch 1, maximum biases were acceptable across all conditions of sample size, maximum inaccuracies were 23.5% for the smallest sample size and decreased with increasing sample size to 5.0% for $N = 15$ and $M = 100$; coverage rates were acceptable across all conditions. For branch 2, high biases and inaccuracies and some slightly too large or too small coverage rates were observed for small sample sizes. These values were observed for σ^2_α which had a rather low true value to begin with in branch 2 ($\sigma^2_\alpha = 0.041$) so that large RB and RI values were to be expected. In all cases in which a too restrictive model (with regard to

Table 5. Most extreme relative biases, maximum relative inaccuracies, and ranges of coverage rates for estimated models that are either more general or equal to the true model in both branches of the simulation study.

	Branch 1			Branch 2		
$N \times M$	m.e. RB	max. RI	CR-range	m.e. RB	max. RI	CR-range
5×25	0.019	0.235	0.925–0.961	0.894	1.057	0.909–1.000
5×50	−0.012	0.162	0.933–0.957	0.534	0.696	0.935–0.991
5×100	0.004	0.119	0.939–0.961	0.304	0.480	0.941–0.992
8×25	−0.008	0.160	0.935–0.961	0.301	0.456	0.936–0.986
8×50	−0.005	0.108	0.935–0.963	0.130	0.324	0.933–0.977
8×100	−0.004	0.080	0.936–0.961	0.042	0.242	0.939–0.970
10×25	−0.006	0.135	0.935–0.961	0.132	0.346	0.936–0.979
10×50	−0.005	0.092	0.934–0.959	0.064	0.265	0.931–0.966
10×100	0.004	0.068	0.931–0.970	0.015	0.186	0.933–0.961
15×25	−0.004	0.101	0.935–0.960	0.029	0.226	0.942–0.960
15×50	−0.005	0.069	0.936–0.957	0.015	0.166	0.933–0.956
15×100	0.003	0.050	0.935–0.959	0.020	0.113	0.936–0.962

Notes. $N \times M$ = sample size comprising N members across M round-robin groups, m.e. RB = most extreme relative bias, max. RI = maximum relative inaccuracy, CR = coverage rate. Very small model parameters (true parameter value $\leq$ 0.01; which applies to $\sigma_{\alpha\beta}$, $\sigma_{\xi_1\zeta_1}$, $\sigma_{\xi_2\zeta_2}$, $\sigma_{\xi_3\zeta_3}$, and $\sigma_{\varepsilon_3\varepsilon_3}$ from branch 2) were not included since RB and RI are always deceptively high for such parameters. See supplementary material for complete results.

the true model) was specified, the sample size did not have a substantial impact on the range of relative biases. In these cases, negative bias ranged from −24.4% ($N = 15$, $M = 25$) to −23.9% ($N = 8$, $M = 50$) in branch 1 and −28.7% ($N = 8$, $M = 50$) to −11.0% ($N = 8$, $M = 25$) in branch 2; whereas positive biases ranged from 38.0% ($N = 8$, $M = 25$) to 38.8% ($N = 15$, $M = 100$) in branch 1 and 30.3% ($N = 10$, $M = 100$) to 43.2% ($N = 5$, $M = 25$) in branch 2.

Concerning the reciprocity correlations, in branch 1, biases larger than 0.05 (or smaller than −0.05) and inaccuracies larger than 0.2 only occurred when the strong invariant model was estimated but not the true model (independent of sample size). Only slight biases and inaccuracies occurred when the weak invariant model was estimated and the strong invariant model was true. In branch 2, reciprocity correlation biases and inaccuracies of this magnitude occurred only when $N = 5$ and $M = 25$ or $N = 5$ and $M = 50$—independent of the combination of true and estimated model. This indicates that, when the variance of the latent actor-effect is small, substantial biases and inaccuracies of reciprocity correlations are expected due to small sample sizes but not to model-misspecifications (see supplementary material for more details).

8.4. Relative Variance Parameters

The supplementary material contains the complete results for the relative variances as digital tables as well as large graphical presentations of biases and inaccuracies for all relative variance parameters. Across both branches, as long as the correct model or a more general was estimated, biases and inaccuracies did not differ much across the different indicators and the sum-score. Table 6 shows ranges of bias and inaccuracy for the sum score in such cases across the different conditions of sample size for both branches. Even with the smallest sample size, biases were below 1 percentage-point for a relative variance parameter in branch 1 and below 3 percentage-points in branch 2. The maximum inaccuracy amounted to 5.5 percentage points for the smallest sample size in branch 2, but was reduced below 1 percentage point for the sample sizes of 8 × 50, 8 × 100, and 15 × 100.

Table 6. Ranges of bias and inaccuracy for sum-score relative variances in cases where the correct model or a more general one was specified across both branches of the simulation study and the different conditions of sample size.

	Branch 1		Branch 2	
$N \times M$	Bias	Inaccuracy	Bias	Inaccuracy
5 × 25	−0.003–0.006	0.036–0.042	−0.024–0.027	0.025–0.055
5 × 50	−0.003–0.004	0.025–0.030	−0.010–0.015	0.019–0.037
5 × 100	−0.006–0.002	0.017–0.021	−0.004–0.006	0.013–0.026
8 × 25	−0.002–0.003	0.022–0.027	−0.005–0.006	0.012–0.028
8 × 50	−0.001–0.001	0.015–0.019	−0.002–0.002	0.009–0.019
8 × 100	−0.001–0.001	0.011–0.014	−0.001–0.001	0.006–0.014
10 × 25	−0.002–0.002	0.018–0.022	−0.003–0.002	0.009–0.023
10 × 50	−0.001–0.002	0.013–0.016	−0.001–0.002	0.006–0.016
10 × 100	−0.001–0.000	0.009–0.011	−0.001–0.001	0.005–0.011
15 × 25	−0.001–0.001	0.013–0.017	−0.002–0.002	0.006–0.016
15 × 50	−0.001–0.001	0.010–0.012	−0.001–0.001	0.004–0.011
15 × 100	−0.000–0.000	0.007–0.008	0.001–0.001	0.003–0.008

Notes. $N \times M$ = sample size comprising N members across M round-robin groups.

Table 7 shows the most extreme biases and inaccuracies observed for relative variance parameters for all indicators and the sum score across all sample sizes for cases in which a too restrictive model was estimated. For branch 1, in all cases in which a strong invariant model was falsely estimated, the most extreme biases and inaccuracies ranged between 5.1 percentage points and 7.0 percentage points. Sample size did not have any influence on biases and only a small influence on inaccuracies and they were no better or worse for the sum-score compared to the individual indicators. A similar pattern is found for the indicators, when the weak invariant model was estimated but the strong model was true, albeit the values were different (maximum bias-range for the first indicator: −0.076 to −0.078, maximum inaccuracy-range for the second indicator: 0.076 to 0.080, maximum bias-range for the second and third indicator: −0.026 to −.028, maximum inaccuracy-range for the first indicator: 0.026 to 0.045). For the sum-score in this misspecification case, biases and inaccuracies were comparatively low to begin with and reduced with increasing sample size (most extreme bias with 5 × 25: −0.007, most extreme inaccuracy with 5 × 25: 0.040, most extreme bias with 15 × 100: −0.005, most extreme inaccuracy with 15 × 100: 0.009). Figure 2 presents relative variance biases in branch 1 across the sample sizes 5 × 25, 5 × 50, and 5 × 100. It shows clearly that, given the configural invariant model is the true model, an unbiased estimation of relative variances for single indicators requires correct model specification. A weak invariant model, however, is equally capable of estimating sum-score relative variances without severe biases. Falsely specifying the strong invariant model (regardless whether the configural or the weak invariant model is true) estimates all relative variances with substantial bias. Increasing the number of round-robin groups has barely any effect on these observations. For branch 2, the concrete values differed to a slight degree but the same general conclusions can be drawn (see supplementary material).

9. Discussion

In this contribution, we presented the different types of measurement invariant social relations—confirmatory factor analysis models, showed how meaningful relative variances of assimilation, consensus and uniqueness can be computed within each of them and executed a simulation study that investigated the impact of (mis-)specifying these models on yielding valid parameter estimates and unbiased and accurate relative variances. When the number of sampled group-members and person-level variances are low, estimation issues for the configural invariant model occur frequently. These can be avoided to a certain degree when weak measurement invariance for the latent variables is implemented (to an even larger degree when strong measurement invariance is implemented). Further, the results suggested that estimating a weak invariant model will not substantially bias the relative variances for the sum-score and reciprocity-

Table 7. Most extreme biases and inaccuracies (in parenthesis) for relative variance parameters of the indicators and sum-scores in too restrictive models across the different conditions of sample size for both branches of the simulation study.

Branch 1

	tcew			tces			twes
N × *M*	*l* = 1	*l* = 2/*l* = 3	*s*	*l* = 1	*l* = 2/*l* = 3	*s*	*l* = 1/*l* = 2/*l* = 3/*s*
5 × 25	−0.078 (0.080)	−0.026 (0.045)	−0.007 (0.040)	−0.056 (0.102)	−0.051 (0.059)	−0.050 (0.059)	−0.063 (0.070)
5 × 50	−0.078 (0.078)	−0.026 (0.036)	−0.007 (0.029)	−0.056 (0.102)	−0.051 (0.053)	−0.051 (0.053)	−0.066 (0.070)
5 × 100	−0.077 (0.077)	−0.026 (0.030)	−0.006 (0.020)	−0.055 (0.102)	−0.052 (0.052)	−0.052 (0.052)	−0.068 (0.068)
8 × 25	−0.076 (0.077)	−0.028 (0.035)	−0.005 (0.026)	−0.054 (0.101)	−0.052 (0.053)	−0.052 (0.052)	−0.066 (0.066)
8 × 50	−0.077 (0.077)	−0.026 (0.029)	−0.005 (0.018)	−0.054 (0.102)	−0.053 (0.053)	−0.053 (0.053)	−0.067 (0.067)
8 × 100	−0.076 (0.076)	−0.026 (0.027)	−0.005 (0.013)	−0.054 (0.102)	−0.054 (0.056)	−0.053 (0.053)	−0.068 (0.068)
10 × 25	−0.076 (0.076)	−0.028 (0.032)	−0.005 (0.026)	−0.054 (0.101)	−0.052 (0.052)	−0.052 (0.052)	−0.067 (0.067)
10 × 50	−0.076 (0.076)	−0.027 (0.028)	−0.005 (0.016)	−0.053 (0.102)	−0.053 (0.053)	−0.053 (0.053)	−0.067 (0.067)
10 × 100	−0.076 (0.076)	−0.026 (0.026)	−0.005 (0.011)	−0.053 (0.103)	−0.054 (0.054)	−0.054 (0.054)	−0.068 (0.068)
15 × 25	−0.077 (0.077)	−0.027 (0.029)	−0.006 (0.017)	−0.054 (0.102)	−0.053 (0.053)	−0.052 (0.052)	−0.067 (0.067)
15 × 50	−0.076 (0.076)	−0.027 (0.027)	−0.005 (0.012)	−0.053 (0.102)	−0.053 (0.053)	−0.053 (0.053)	−0.068 (0.068)
15 × 100	−0.077 (0.077)	−0.026 (0.026)	−0.005 (0.009)	−0.054 (0.103)	−0.054 (0.054)	−0.054 (0.054)	−0.068 (0.068)

Branch 2

	tcew			tces			twes
N × *M*	*l* = 1	*l* = 2/*l* = 3	*s*	*l* = 1	*l* = 2/*l* = 3	*s*	*l* = 1/*l* = 2/*l* = 3/*s*
5 × 25	0.051 (0.060)	−0.022 (0.045)	0.010 (0.045)	−0.062 (0.070)	0.020 (0.045)	−0.022 (0.049)	−0.020 (0.051)
5 × 50	0.054 (0.057)	−0.018 (0.033)	−0.005 (0.032)	0.060 (0.062)	0.013 (0.032)	−0.017 (0.035)	−0.018 (0.038)
5 × 100	0.055 (0.056)	−0.017 (0.025)	−0.003 (0.023)	0.061 (0.061)	0.012 (0.023)	−0.016 (0.027)	−0.020 (0.030)
8 × 25	0.053 (0.054)	−0.020 (0.030)	0.001 (0.026)	0.059 (0.060)	0.013 (0.027)	−0.016 (0.029)	−0.016 (0.030)
8 × 50	0.054 (0.054)	−0.019 (0.024)	−0.001 (0.020)	0.061 (0.061)	0.012 (0.020)	−0.016 (0.024)	−0.018 (0.024)
8 × 100	0.055 (0.055)	−0.018 (0.020)	−0.002 (0.014)	0.061 (0.061)	0.012 (0.016)	−0.017 (0.020)	−0.019 (0.021)
10 × 25	0.052 (0.053)	−0.020 (0.028)	0.001 (0.023)	0.059 (0.059)	−0.013 (0.025)	−0.015 (0.026)	−0.017 (0.026)
10 × 50	0.054 (0.054)	−0.019 (0.022)	−0.001 (0.015)	0.061 (0.061)	0.013 (0.018)	−0.017 (0.021)	−0.018 (0.021)
10 × 100	0.055 (0.055)	−0.018 (0.019)	−0.002 (0.011)	0.062 (0.062)	0.012 (0.014)	−0.018 (0.019)	−0.019 (0.020)
15 × 25	0.053 (0.053)	−0.019 (0.023)	0.001 (0.016)	0.060 (0.060)	0.012 (0.019)	−0.016 (0.021)	−0.019 (0.022)
15 × 50	0.053 (0.053)	−0.019 (0.020)	−0.001 (0.011)	0.061 (0.061)	0.012 (0.015)	−0.017 (0.019)	−0.019 (0.020)
15 × 100	0.054 (0.054)	−0.018 (0.018)	−0.002 (0.008)	0.062 (0.062)	0.012 (0.012)	−0.018 (0.018)	−0.019 (0.019)

Notes. tcew = the weak model was estimated when the configural model was true, tces = the strong model was estimated when the configural model was true, twes = the strong model was estimated when the weak model was true, *N* × *M* = sample size comprising *N* members across *M* round-robin groups. *l* marks an indicator, *s* reflects the sum score. In the true configural invariant model, factor loadings are identical across the second and third indicator, so biases and inaccuracies are also identical. In the true weak invariant model, factor loadings are identical across all indicators, so biases and inaccuracies are identical across all indicators and the sum-score.

correlations, even when only configural invariance holds in the true model. However, mis-specifying a model as strong invariant will substantially bias all absolute and relative parameters. Note that detecting a strong invariant model as too restrictive via a likelihood-ratio test worked almost perfectly in the simulation study (Table 4).

9.1. Recommendations

The most crucial finding of the simulation study is the robustness of relative variance parameters for the sum-score against mis-specifying a configural invariant SR-CFA as a weak invariant SR-CFA. Moreover, it can also be regarded a fortunate finding that reciprocity-correlations are robust in such cases. However, one needs to keep in mind that the results of any Monte-Carlo simulation cannot be generalized to conditions that were not under investigation—we do not know, for example, if this robustness still stands when true parameter-values are different to those chosen for the current contribution. On this basis, we make the following recommendations to applied researchers particularly interested in the relative variance components:

1. As in any type of invariance-analysis, the configural invariant model should be estimated first. Stricter forms of measurement invariance should be implemented step by step and the strictest type of measurement invariance that does not substantially deter model fit according to the likelihood-ratio tests should be maintained.
2. One should always investigate the effect of measurement invariance on the relative variance components and reciprocity correlations. Conclusions are corroborated if this impact is only small (e.g., if the rank-order of the components with regard to the relative amount of variance explained remains unaffected).
3. If a more general type of invariance leads to inadmissible solutions, one should take the estimated parameters of a more restrictive model, add a range of plausible values for the inadmissible parameters of the more general model, and conduct a simulation study to investigate the effect of measurement invariance specification on the relative variance components.

With regards to 3., we note that the srm-package by Nestler et al. (2020) not only enables specifying SR-CFAs in a convenient way similar to the lavaan-syntax (Rosseel, 2012) but also conducting small simulation studies very efficiently (see the supplementary materials of Nestler et al., 2020). However, if models become more complex (e.g., the SR-CFA contains multiple constructs each assessed by multiple indicators) the less user-friendly wide-data format

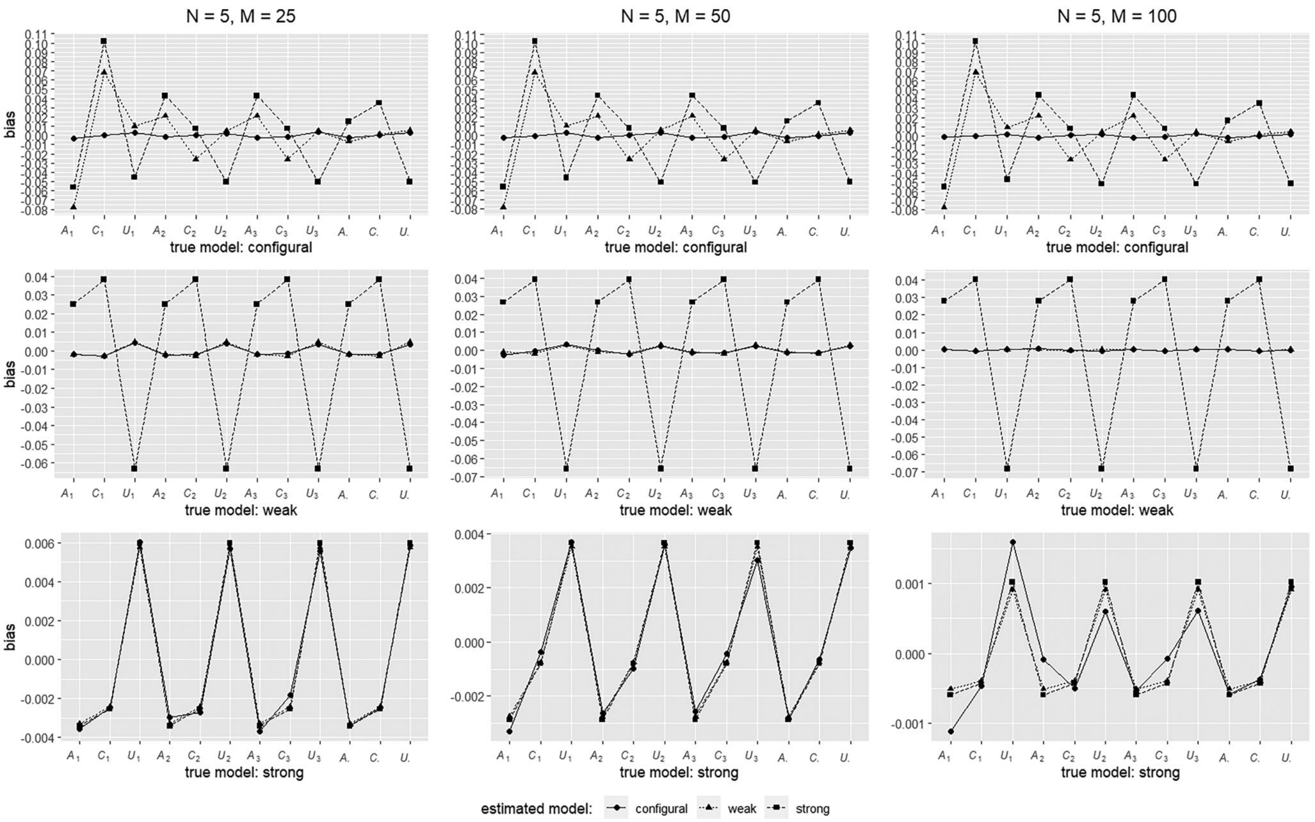

Figure 2. Biases for relative variance parameters in branch 1 of the simulation study across three selected conditions of sample size. N = number of round-robin group members, M = number of round-robin groups, *A* = assimilation, *C* = consensus, *U* = uniqueness. There were three different indicators (*l* = 1, 2, or 3). A dot after a letter indicates the relative variance for the sum-score.

approach (Jendryczko & Nussbeck, 2024) might be unavoidable.

9.2. Substantive Deliberations and Future Research

From a substantive perspective, researchers may want to aim to create measurement instruments for which at least weak measurement invariance holds—even with the problem of estimation issues set aside. The variance decomposition of a sum-score is usually less of interest; researchers will mainly want to compute assimilation, consensus, and uniqueness for the variances given by the latent variables the measurement instrument is supposed to assess (the latent common actor- partner- and relationship-effects). However, as we have shown, directly comparing the variances of these is only valid when at least weak measurement invariance holds as otherwise they are measured on different scales. This means that differences between persons do not manifest to the same degree as differences between relationships in the observations. However, if weak (or strong) invariance does not hold, less restrictive alterations of a configural invariant SR-CFA may also be computed. Eid et al. (2024) discuss SR-CFAs in which the random intercepts on person-level are discarded for only one indicator. These models are then consistent with stochastic measurement and classical test theory, which gives all variables are clearer meaning based on conditional expectations (which is only the case for the strong invariant model of all models from this contribution; see Jendryczko & Nussbeck, 2024b). In these models, the different indicators for a construct should be regarded different methods for assessing the construct, rendering a variance decomposition in the framework of multitrait-multimethod analysis meaningful (Eid et al., 2003; Jendryczko & Nussbeck, 2024a; Koch et al., 2016). Generally, while we focused on classical conceptualizations of assimilation, consensus, and uniqueness in this contribution, other variance decompositions are meaningful and informative. Bonito & Kenny (2010) discuss the difference of "stable" SRM-variance components (represented here by σ^2_α, σ^2_β, and σ^2_γ) manifesting in all indicators and "unstable" indicator-specific components (represented here by $\sigma^2_{\xi_l}$, $\sigma^2_{\zeta_l}$, and $\sigma^2_{\varepsilon_l}$). In order to apply their formulas, however, all factor loadings must be restricted to 1. If that assumption is too restrictive, one needs to focus on the sum-score again and not assess the latent variables directly. In any case, researchers are well advised to also consider the remaining variance components given by $\sigma^2_{\xi_l}$, $\sigma^2_{\zeta_l}$, and $\sigma^2_{\varepsilon_l}$ in order to evaluate the homogeneity of their measurement instrument. Also, as Jendryczko (2022) illustrated, absolute and relative variances should always be investigated alongside each other.

In this context, future work is also needed to evaluate fit for SR-CFAs. In this contribution, we focused on standard likelihood-ratio tests in order to detect violations of measurement invariance. Future work should also establish alternative strategies for evaluating fit of SR-CFAs both in the context of assessing measurement invariance but also—more broadly—for establishing strategies for investigating the number of assessed constructs (for first discussions on this, see Jendryczko & Nussbeck, 2024, 2024a). Relatedly, it is important to note that we focused on a single-construct SR-CFA with specific population parameters and specific distributional assumptions. Future work will also need to investigate the generalizability of our findings to multiple construct SR-CFAs with different population parameters and distributions. This will also help to establish benchmarks for judging the severity of absolute biases of relative variance parameters.

Lastly, other estimation techniques for SR-CFAs should be established. In the multilevel context, it has been shown that the problem of inadmissible solutions with small sample sizes and small true variances can be effectively avoided with Bayesian estimation (e.g., Depaoli & Clifton, 2015). Bayesian structural equation modeling for round-robin data is still in rather early steps of development. Lüdtke et al. (2013) presented a Bayesian framework for the estimation of the standard (single-indicator) SRM which displayed favorable statistical properties in comparison to the traditional ANOVA-based approach. While they stated that the framework can be expanded to measurement models, this is yet to be explicitly shown and evaluated. Bhangale and Jorgensen (2024; see also: Jorgensen et al., 2024) compared Bayesian and maximum likelihood estimation of the multivariate SRM (containing multiple constructs each measured with one indicator and, thus, no measurement models) in a first small simulation study and found the Bayesian approach to be inferior. It also remains to be seen how Bayesian model comparison techniques like the Bayes-factor and the Posterior Predictive Checking procedure (e.g., Depaoli et al., 2023) can handle the comparison of different types of measurement invariance in SR-CFA. The interpretational issues of latent variables when only configural measurement invariance holds are, of course, present independently from the estimation technique.

References

Bhangale, A. M., & Jorgensen, T. D. (2024). Comparing maximum likelihood to Markov chain Monte Carlo estimation of the multivariate social relations model. In M. Wiberg, J.-S. Kim, H. Hwang, H. Wu, & T. Sweet (Eds.), *Quantitative psychology: The 88th Annual Meeting of the Psychometric Society, Maryland, USA, 2023* (Vol. 452, pp. 65–75). Springer Proceedings in Mathematics & Statistics. Springer. https://doi.org/10.1007/978-3-031-55548-0_7

Bonito, J. A., & Kenny, D. A. (2010). The measurement of reliability of social relations components from round-robin designs. *Personal Relationships*, *17*, 235–251. https://doi.org/10.1111/j.1475-6811.2010.01274.x

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling*, *22*, 327–351. https://doi.org/10.1080/10705511.2014.937849

Depaoli, S., Kaplan, D., & Winter, S. D. (2023). Foundations and extensions in Bayesian structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed., pp. 701–721). The Guilford Press.

Eid, M., Geiser, C., & Koch, T. (2024). *Structural equation modeling of multiple rater data*. Guilford Publications.

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple- indicator CT-C(M-1) model. *Psychological Methods*, *8*, 38–60. https://doi.org/10.1037/1082-989x.8.1.38

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2020). *mvtnorm: Multivariate Normal and t Distributions. R package version 1.0-12.* http://CRAN.R-project.org/package=mvtnorm

González-Romá, V., & Hernández, A. (2017). Multilevel modeling: Research-based lessons for substantive researchers. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 183–210. https://doi.org/10.1146/annurev-orgpsych-041015-062407

Jak, S. (2019). Cross-level invariance in multilevel factor models. *Structural Equation Modeling*, *26*, 607–622. https://doi.org/10.1080/10705511.2018.1534205

Jak, S., Oort, F. J., & Dolan, C. V. (2014). Using two-level factor analysis to test for cluster bias in ordinal data. *Multivariate Behavioral Research*, *49*, 544–553. https://doi.org/10.1080/00273171.2014.947353

Jendryczko, D. (2022). Multiple group structural equation modeling of the social relations model. *Structural Equation Modeling*, *29*, 791–799. https://doi.org/10.1080/10705511.2022.2032715

Jendryczko, D., & Nussbeck, F. W. (2024). Estimating and investigating multiple constructs multiple indicators social relations models with and without roles within the traditional structural equation modeling framework: A tutorial. *Psychological Methods*, *29*, 919–946. https://doi.org/10.1037/met0000534

Jendryczko, D., & Nussbeck, F. W. (2024a). A correlated traits correlated (methods – 1) multitrait-multimethod model for augmented round-robin data. *British Journal of Mathematical and Statistical Psychology*, *77*, 1–30. https://doi.org/10.1111/bmsp.12324

Jendryczko, D., & Nussbeck, F. W. (2024b). Latent reciprocal engagement and accuracy variables in social relations structural equation modeling. *Multivariate Behavioral Research*, *60*, 115–137. Advance online publication. https://doi.org/10.1080/00273171.2024.2386060

Jorgensen, T. D., Bhangale, A. M., & Rosseel, Y. (2024). Two-stage limited-information estimation for structural equation models of round-robin variables. *Stats*, *7*, 235–268. https://doi.org/10.3390/stats7010015

Kenny, D. A. (2020). *Interpersonal perception: The foundation of social relationships* (2nd ed.). Guilford Press.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. Guilford Press.

Kerkhoff, D., & Nussbeck, F. W. (2022). Obtaining sound intraclass correlation and variance estimates in three-level models: The role of sampling-strategies. *Methodology*, *18*, 5–23. https://doi.org/10.5964/meth.7265

Kim, E. S., & Cao, C. (2015). Testing group mean differences of latent variables in multilevel data using multiple-group multilevel CFA and multilevel MIMIC modeling. *Multivariate Behavioral Research*, *50*, 436–456. https://doi.org/10.1080/00273171.2015.1021447

Koch, T., Schultze, M., Jeon, M., Nussbeck, F. W., Praetorius, A.-K., & Eid, M. (2016). A cross-classified CFA-MTMM model for structurally different and nonindependent interchangeable methods. *Multivariate Behavioral Research*, *51*, 67–85. https://doi.org/10.1080/00273171.2015.1101367

Li, X., & Beretvas, S. N. (2013). Sample size limits for estimating upper level mediation models using multilevel SEM. *Structural Equation Modeling*, *20*, 241–264. https://doi.org/10.1080/10705511.2013.769391

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy–Bias trade-offs in full and partial error correction models. *Psychological Methods*, *16*, 444–467. https://doi.org/10.1037/a0024376

Lüdtke, O., Robitzsch, A., Kenny, D. A., & Trautwein, U. (2013). A general and flexible approach to estimating the social relations model using Bayesian methods. *Psychological Methods*, *18*, 101–119. https://doi.org/10.1037/a0029252

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

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*, 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*, 535–549. https://doi.org/10.1007/s11336-014-9435-8

Nestler, S., Lüdtke, O., & Robitzsch, A. (2020). Maximum likelihood estimation of a social relations structural equation model. *Psychometrika*, *85*, 870–889. https://doi.org/10.1007/s11336-020-09728-z

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

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

Salazar Kämpf, M., Liebermann, H., Kerschreiter, R., Krause, S., Nestler, S., & Schmukle, S. C. (2018). Disentangling the sources of mimicry: Social relations analyses of the link between mimicry and liking. *Psychological Science*, *29*, 131–138. https://doi.org/10.1177/0956797617727121

Warner, R. M., Kenny, D. A., & Stoto, M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology*, *37*, 1742–1757. https://doi.org/10.1037/0022-3514.37.10.1742

Appendix A

Illustrating the Simplification of Equation (12) for a Weak or Strong Invariant Model

Consider the computation of assimilation for the sum-score within a weak or strong invariant model with only two indicators ($\lambda_{\alpha_1} = \lambda_{\beta_1} = \lambda_{\gamma_1} = \lambda_1$ and $\lambda_{\alpha_2} = \lambda_{\beta_2} = \lambda_{\gamma_2} = \lambda_2$). Equation (12) simplifies to Equation (13) in the following way:

$$\begin{aligned} A_\bullet &= \frac{\lambda_1^2\sigma_\alpha^2 + \lambda_2^2\sigma_\alpha^2 + 2\lambda_1\lambda_2\sigma_\alpha^2}{(\lambda_1^2\sigma_\alpha^2 + \lambda_2^2\sigma_\alpha^2 + 2\lambda_1\lambda_2\sigma_\alpha^2) + (\lambda_1^2\sigma_\beta^2 + \lambda_2^2\sigma_\beta^2 + 2\lambda_1\lambda_2\sigma_\beta^2) + (\lambda_1^2\sigma_\gamma^2 + \lambda_2^2\sigma_\gamma^2 + 2\lambda_1\lambda_2\sigma_\gamma^2)} \\ &= \frac{\sigma_\alpha^2(\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2)}{\sigma_\alpha^2(\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2) + \sigma_\beta^2(\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2) + \sigma_\gamma^2(\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2)} \\ &= \frac{\sigma_\alpha^2(\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2)}{(\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2)(\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2)} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2}. \end{aligned} \tag{A1}$$