Erschienen in: Structural Equation Modeling : A Multidisciplinary Journal ; 29 (2022), 5. - S. 791-799
https://dx.doi.org/10.1080/10705511.2022.2032715

Multiple Group Structural Equation Modeling of the Social Relations Model

David Jendryczko

University of Konstanz

ABSTRACT

The social relations model is a statistical tool that allows the analysis of group dynamics as dyadic interactions between individuals. Within a multiple group structural equation modeling framework, Wald-tests and likelihood ratio tests based on (1) equality constraints among model parameters and (2) Lagrange multipliers for restrictions among non-linear parameter transformations are presented as methods for group-comparisons of various quantifications of group dynamics that hold different interpretations. The methods are illustrated with an empirical example. A simulation study investigates the performance of the methods with regard to Type I error-rate recoverability and statistical power and displays overall promising results. Implications and limitations are discussed.

KEYWORDS
Dyadic data; group dynamics; Lagrange-multiplier; multiple groups; social relations model

1. Introduction

The social relations model (SRM; Kenny et al., 2006; Warner et al., 1979) is a conceptual and statistical tool that enables the analysis of dyadic interactions between individuals in groups. It was originally developed to decompose directed behavior within social dyads (e.g., person A rates how much he or she likes person B) into variance components on person-level (Does A like B because A is a general "liker" and/or because B is generally likeable?) and dyad level (Does A like B because B is "right up A's ally"?). These variance components suffice, when the analysis is employed on only one group or on several groups that are assumed to share the same SRM parameters (i.e., stem from the same population). One example for the latter case are zero-acquaintance studies where study participants are randomly assigned to "artificial" groups in which they rate each other (e.g., Salazar Kämpf et al., 2018). Due to the randomization process, all observed group differences in average ratings and other SRM-parameters are assumed to emerge from sampling error.

In a lot of cases, however, groups are not created by the experimental design but are culturally and/or naturally fixed (e.g., school-classes, families, workgroups, or therapy-groups; Betts et al., 2012; Buist et al., 2008; Markin & Kivlighan, 2008; Sfetcu, 2013). Additionally, even groups of the same "type" might be structurally different: different school classes encapsulate different age ranges, different families have different socio-economic statuses, different workgroups stem from different market enterprises, and different therapy groups operate on different therapeutic backgrounds. This ascertainment highlights the importance of the SRM for different psychological disciplines, such as social, personality-, educational, developmental, organizational, and clinical psychology, and for its use for the study of group dynamics, which authors have encouraged explicitly (Christensen & Feeney, 2016; Kenny et al., 2015; Marcus, 1998, 2021). For example, the SRM can be used to investigate whether negativity between an adolescent child with externalizing problem behavior and his or her family members generalizes to other dyadic interactions (e.g., between the mother and the father) within the family (Eichelsheim et al., 2011).

However, the SRM has received rather sparse attention for the study of *differences in group-dynamics between groups* (but see Eichelsheim et al., 2011; Loeys et al., 2021; Stas et al., 2015). Ervin and Bonito (2014) executed a detailed review of applications of the SRM. Out of the 107 SRM studies they reviewed, more than 70% pooled all investigated groups for the estimation of the SRM parameters. Most of the remaining studies estimated separate parameters for separate types of groups (for example, students and workers; Lam et al., 2011) but did not further investigate the differences between groups.

The object of the current contribution is two-fold. Firstly, we present how differences of the SRM parameters between groups can be analyzed and tested within a multiple groups structural equation modeling (mgSEM) framework. Illustrations are complemented by exemplary executions of the analyses in R in the supplemental material. Secondly, we present the results of a simulation study aiming at determining the minimum required number of groups and group sizes for reliably detecting quantitatively varying group differences of the SRM parameters. In the following, we will first recap the structural equation modeling of the SRM.

CONTACT David Jendryczko david.jendryczko@uni-konstanz.de University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany.

2. The Social Relations Model and Its Formulation within SEM

An SRM-analysis is executed with (usually completely balanced) round-robin data. When the behavior of every individual toward every other individual within a group (e.g., a liking-rating; see above) is assessed, then the resulting variation across the judgments constitutes a round-robin variable. The basic model equation (Kenny, 2020; Warner et al., 1979) for this variable is:

$$X_{ijk} = \mu_k + \alpha_{ik} + \beta_{jk} + \gamma_{ijk} \text{ with } i \neq j. \tag{1}$$

X_{ijk} represents the behavior of actor i toward partner j in group k. For simplicity, we will abide by the example of a liking-rating for the rest of the explanation. μ_k depicts the average liking rating in group k. Group-level effects can be omitted from the data, for example, by group-mean-centering (Nestler, 2016). α_{ik} represents the actor-effect of person i in group k and describes the extent to which person i tends to generally like other people. β_{jk} represents the partner-effect of person j in group k and depicts the extent to which person j tends to be generally liked by other people. These two parameters are person-level effects. Finally, γ_{ijk} is the relationship-effect of actor i toward partner j in group k and depicts i's unique liking of j (dyad-level effect). Note that a relationship is "directed." In other words, a dyad containing the two persons i and j holds two juxtaposed relationships: ij and ji.

On the group level, μ_k is assumed to be normally distributed with an expectancy of $\bar{\mu}$ (the grand-mean across all groups) and variance of σ^2_μ. On the person-level, the actor- and partner-effect for every person i is assumed to follow a bivariate normal distribution:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim MVN(0, \Sigma_p) \text{ with } \Sigma_p = \begin{pmatrix} \sigma^2_\alpha & \sigma_{\alpha\beta} \\ & \sigma^2_\beta \end{pmatrix}. \tag{2}$$

σ^2_α and σ^2_β depict the actor-effect variance and the partner-effect variance, respectively. The actor-partner-covariance $\sigma_{\alpha\beta}$ denotes generalized reciprocity (Are people that tend to like others also generally liked by others?). On the dyad-level, the two relationship effects for every dyad ij are also assumed to follow a bivariate normal distribution:

$$\begin{pmatrix} \gamma_{ij} \\ \gamma_{ji} \end{pmatrix} \sim MVN(0, \Sigma_d) \text{ with } \Sigma_d = \begin{pmatrix} \sigma^2_\gamma & \sigma_{\gamma\gamma} \\ & \sigma^2_\gamma \end{pmatrix}. \tag{3}$$

Here, σ^2_γ depicts the relationship-effect variance and the relationship covariance $\sigma_{\gamma\gamma}$ denotes dyadic reciprocity (Is a person that likes a specific different person also more likely to be liked back by him or her?). It follows that the variance σ^2_X across the round-robin judgments is decomposed into four orthogonal components:

$$\sigma^2_X = \sigma^2_\mu + \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\gamma. \tag{4}$$

As Equation (1) shows, an SRM can be conceptualized as a special case of confirmatory factor analysis (CFA) in which every observation is explained by a set of four latent variables. For an SRM-analysis in the classic SEM framework, the round-robin judgments must be structured in a wide data format. That is, each row depicts a different group k and every column contains the round-robin judgment of a specific actor i on a specific actor j (Kenny, 2020; Kenny & Livi, 2009). If, for example, our dataset contains four individuals per group (the minimum required group size to identify the SRM; Warner et al., 1979) then we would obtain twelve columns: X_{12}, X_{13}, X_{14}, X_{21}, X_{23}, X_{24}, X_{31}, X_{32}, X_{34}, X_{41}, X_{42}, and X_{43}. The assignment of a specific person to a specific index (1, 2, 3, or 4 in this case) is arbitrary. Equation (1) is then employed on every column. This implies that we must specify every person-level effect and every actor-partner covariance for every person-index, every relationship-effect for every relationship index, and every relationship covariance for every pair of juxtaposed relationship indices. Afterward, we must impose equality constraints on the variance- and covariance-parameters across persons and relationships, respectively. The equality restrictions account for interchangeable dyads in which the group-members do not hold specific roles (e.g., equal team members; Oosterhof, et al., 2009) as opposed to non-interchangeable dyads (e.g., in families; Eichelsheim et al., 2011). In the current contribution, we focus on mgSEM of the SRM with interchangeable dyads (see Figure 1 for an SEM SRM with interchangeable dyads and four persons per group), but mgSEM of the SRM with non-interchangeable dyads is also possible and straightforwardly applied (see, for example, Stas et al., 2015). Note that unequal group sizes create missing values by design that can be handled with the full information maximum likelihood method (FIML, see Table 1). Note further, that the model fit indices are irrelevant because the standard saturated model is not suitable in the case of interchangeable dyads and the SRM itself is, in fact, saturated (Kenny & Livi, 2009; Olsen & Kenny, 2006).

3. Multiple Group Structural Equation Modeling of the SRM

So far, we have used the term "group" generally, but now we must distinguish between a round-robin group (short: rr-group) and an analysis group (short: a-group). An rr-group is, as the name suggests, a group in which a round-robin variable has been assessed. An a-group is a group that is hypothesized to be at least in some aspects structurally different from at least on other a-group, i.e., holds at least partially different population SRM-parameters. In mgSEM, a single rr-group can constitute its own a-group in the case of interchangeable dyads (as long as FIML is used; Nestler et al., 2020; Voelkle et al., 2012), but several rr-groups can also be united to an a-group. Consider the research on personality perception upon zero acquaintance conducted by Albright et al., 1988). Their data ("Zero"; openly accessible at http://davidakenny.net/srm/srmdata.htm and in the supplemental material to this article) contains 36 rr-groups (each containing four to six rr-group members) nested in four different psychology courses, nested in two different colleges. A-groups for substantive comparisons could be constructed based on rr-group, course, or college (see Table 1). For example, if one college is considered to hold a more decorated academic reputation, one might theorize that the students' interpersonal perception of the personality facet

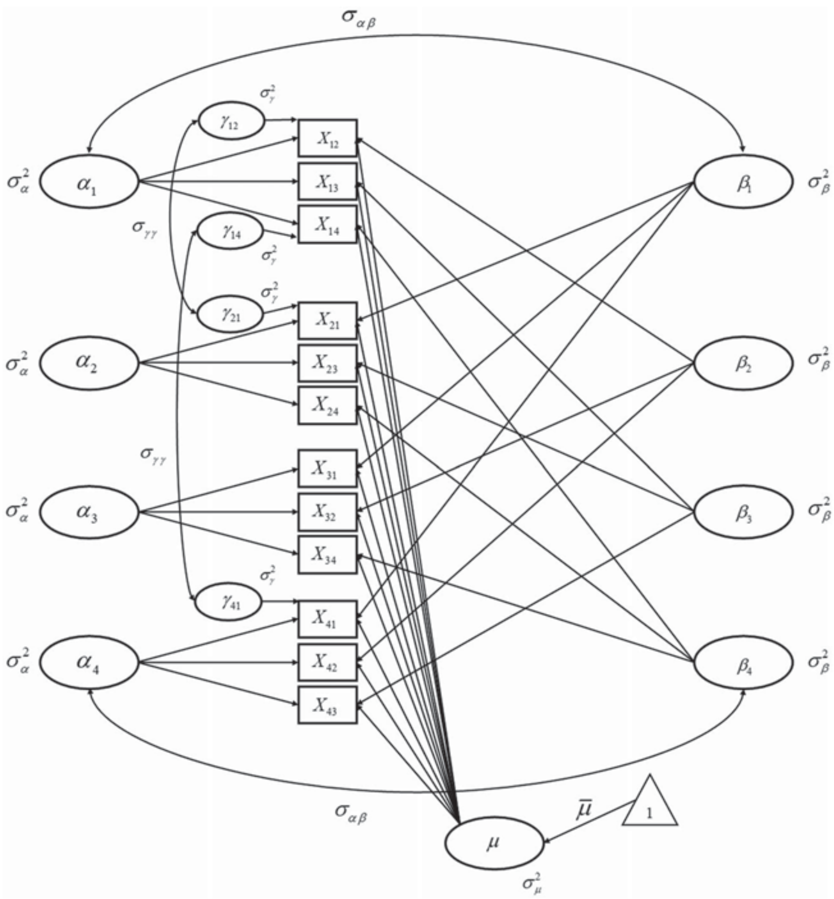

Figure 1. The social relations model with interchangeable dyads and four persons per group as a structural equation model. The actor-effect variance σ^2_α is constrained to equality across all persons as actors, the partner-effect variance σ^2_β is constrained to equality across all persons as partners, the relationship-effect variance σ^2_γ is constrained to equality across all relationships (for simplicity, only four of the twelve are shown), the actor-partner covariance $\sigma_{\alpha\beta}$ is constrained to equality across all persons (for simplicity, only two of the four are shown), and the relationship covariance $\sigma_{\gamma\gamma}$ is constrained to equality across all dyads (for simplicity, only two of the six are shown). All factor loadings are restricted to 1.

Table 1. Data excerpt (wide-format) from Albright et al. (1988) on students' interpersonal perception of intellect upon zero acquaintance.

X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{21}	…	X_{65}	rr-group-ID	Course-ID	College-ID
6	6	7	6	*MV*	5	…	*MV*	1	1	1
5	5	5	*MV*	*MV*	4	…	*MV*	2	2	1
7	7	7	*MV*	*MV*	6	…	*MV*	3	2	1
6	7	5	5	*MV*	5	…	*MV*	4	3	1
6	5	6	6	4	4	…	6	5	4	2

Notes. X_{ij}: intellect-rating of actor *i* on partner *j*; rr-group-ID: round-robin-group-ID; *MV*: Missing Value by Design (rr-groups contained four to six individuals). rr-groups were nested in four different psychology courses which were nested in two different colleges. Substantive analysis-groups can be constructed based on rr-group-ID, course-ID, or college-ID.

"intellect" (DeYoung et al., 2007) operates differently in the colleges. Two a-groups with the first one encapsulating all rr-groups from the first college and the second one encapsulating all rr-groups from the second college would then be constructed for the comparison within the mgSEM framework. We will use this example for our illustrations.[1]

[1]Such a research-question might seem far-fetched. However, we deemed it appropriate for illustration purposes because a real data application is more interesting than the analysis of simulated data and because we can check whether such a weakly constructed hypothesis will actually be met with lack of evidence. More useful examples for the employment of the approach will be given in the remainder.

Comparisons between a-groups can be conducted based on (the grand mean and) absolute values of the SRM-variance parameters as long as the same round-robin measure (i.e., the same metric) has been applied across groups. However, by means of the SRM, the orthogonal variance components on person- and dyad-level are also investigated in terms of their sizes relative to their sum (Kenny, 2020). This procedure provides three different intra-class-correlation coefficients (ICCs) depicting the proportion of the total variance in dyadic interaction that can be attributed to differences in actor-, partner-, and relationship-effects, constituting assimilation (ICC_α), consensus (ICC_β) and uniqueness (ICC_γ), respectively:

$$\begin{aligned} ICC_\alpha &= \sigma^2_\alpha/(\sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\gamma) \\ ICC_\beta &= \sigma^2_\beta/(\sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\gamma) \\ ICC_\gamma &= \sigma^2_\gamma/(\sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\gamma). \end{aligned} \tag{5}$$

If uniqueness exceeds .5, then most of the variance in dyadic interactions can be explained by the unique relationship between actor and partner, as opposed to characteristics grounded in the individual (actor- and partner-effects). If consensus exceeds .5, then most of the variance in dyadic interactions can be explained by specific individual characteristics of the partner. This, in turn, implies that the actor-variance (and relationship-variance) is comparatively low—in other words, actors consent to a high degree in the perception of others. Conversely, if assimilation exceeds .5, then most of the variance is explained by specific individual characteristics of the actors which implies that the partner-variance (and relationship-variance) is comparatively low: actors assimilate their perception-template across different partners to a high degree. Note that only uniqueness derives a specific comparison between dyad-level effects on one hand and person-level effects on the other hand, whereas assimilation and consensus compare a specific person-level effect to the dyad-level effect and the respective juxtaposed person-level effect. It might therefore also be useful to investigate the person-level variance, specifically, to determine whether it is mostly comprised of person-level-exclusive assimilation ($ICC_{\alpha p}$) or person-level exclusive consensus ($ICC_{\beta p}$):

$$\begin{aligned} ICC_{\alpha p} &= \sigma^2_\alpha/(\sigma^2_\alpha + \sigma^2_\beta) = 1 - ICC_{\beta p} \\ ICC_{\beta p} &= \sigma^2_\beta/(\sigma^2_\alpha + \sigma^2_\beta) = 1 - ICC_{\alpha p}. \end{aligned} \tag{6}$$

As Kenny (2020, p. 299) noted, the distinction between absolute and relative variances is crucial for a-group comparisons. Absolute variances can differ across a-groups, while the ICCs are equal and differences in all ICCs can be caused by an absolute difference in only one or two of the three variance components. For example, imagine we compared two different psychotherapy (a-, and, simultaneously, rr-) groups of drug-addicted patients. Group 1 employed standard conversation-based group therapy (control condition) while group 2 (treatment condition) employed wilderness therapy (e.g., Christensen & Feeney, 2016). In clinical psychology, one theoretical conceptualization of drug addiction views it as a form of attachment disorder, hindering the patients from building meaningful relationships (Flores, 2001). Imagine further, that the perceived friendship toward the other group members was assessed several times during the program. In the beginning and middle of the program, both groups displayed negligible partner-effect and relationship-effect variances but moderate actor-effect variances, thereby being mainly characterized by assimilation. Patients differed to a certain degree on whether they viewed different group members as friends but no one was generally seen as a friend, nor did any unique friendships emerge. At the end of the program, the wilderness-therapy group displayed a much larger degree of uniqueness than the control group. Based on this finding, one might be tempted to conclude that the wilderness therapy was more successful in fostering unique friendships among the group members, but it is important to compare absolute variances to understand how the different treatments changed the group dynamics. In one hypothetical scenario, the actor-effect variance simply reduced substantially in only the wilderness group while the (partner-effect and) relationship-effect variance remained constant in both groups. In another scenario, actor- and partner-effect variances decreased and the relationship variance increased substantially in the wilderness group while no changes were observed in the control group. Only the latter scenario would be interpreted as a success of the wilderness therapy in actually fostering unique relationships.[2]

3.1. Testing Group Differences of Absolute Variances

Testing a-group differences of absolute variances (and the grand mean) can be straightforwardly executed with Wald-tests and Likelihood-Ratio (LR) tests involving equality constraints (e.g., Stas et al., 2015). The Wald test compares two estimated SRM-parameters from two different a-groups of an estimated mgSEM directly:

$$W = \frac{\psi_1 - \psi_2}{\sqrt{\sigma^2_{\psi_1} + \sigma^2_{\psi_2} - 2\sigma_{\psi_1\psi_2}}}. \tag{7}$$

Here, ψ_1 and ψ_2 denote the estimates for an SRM-parameter from a-group 1 and 2, respectively. $\sigma^2_{\psi_1}$ and $\sigma^2_{\psi_2}$ denote their respective variances and $\sigma_{\psi_1\psi_2}$ denotes their covariance. W approximately follows a standard normal distribution under the H_0 ($\psi_1 = \psi_2$) and can therefore be used for a z-test.

The LR test requires the estimation of two mgSEMs. In the H_1-model, all SRM-parameters are freely estimated in all a-groups. The H_0-model contains one or several equality constraints among SRM-parameters across a-groups:

$$\theta_1 = \theta_2 = \ldots = \theta_M \tag{8}$$

with θ_m denoting a vector of parameters in a-group m and M denoting the total number of a-groups that are considered for the comparison. Under the H_0, the doubled difference of the log-likelihoods (LL) of the two models

$$LR = 2LL_{H_1} - 2LL_{H_0} \tag{9}$$

approximately follows a χ^2-distribution with the degrees of freedom equaling the number of the equality constraints. Testing one specific absolute SRM-parameter for equality across a-groups requires the exclusive consideration of this parameter, implying a single equality constraint.

3.2. Testing Group Differences of ICCs and Reciprocity-Coefficients

For testing a-group differences of relative variances and reciprocity-coefficients, we suggest using Lagrange-multipliers (e.g., Yuan & Liu, 2021) to augment the mgSEM-fitting-function by equality-constraints among non-linear functions of the SRM-parameters and then executing an LR test. The following equality constraints can be implemented

[2]We note that a difference of SRM-parameters between a control and a treatment-group can only be interpreted as a causal effect resulting from the treatment if certain requirements are met. We will elaborate on this in the discussion.

in the H_0-model to test for equality of ICC_α, ICC_β, ICC_γ, $ICC_{\alpha p}$, and $ICC_{\beta p}$, respectively:

$$\begin{gathered}\sigma^2_{\alpha_1}/(\sigma^2_{\alpha_1}+\sigma^2_{\beta_1}+\sigma^2_{\gamma_1})=\sigma^2_{\alpha_2}/(\sigma^2_{\alpha_2}+\sigma^2_{\beta_2}+\sigma^2_{\gamma_2})=\ldots=\sigma^2_{\alpha_M}/(\sigma^2_{\alpha_M}+\sigma^2_{\beta_M}+\sigma^2_{\gamma_M})\\ \sigma^2_{\beta_1}/(\sigma^2_{\alpha_1}+\sigma^2_{\beta_1}+\sigma^2_{\gamma_1})=\sigma^2_{\beta_2}/(\sigma^2_{\alpha_2}+\sigma^2_{\beta_2}+\sigma^2_{\gamma_2})=\ldots=\sigma^2_{\beta_M}/(\sigma^2_{\alpha_M}+\sigma^2_{\beta_M}+\sigma^2_{\gamma_M})\\ \sigma^2_{\gamma_1}/(\sigma^2_{\alpha_1}+\sigma^2_{\beta_1}+\sigma^2_{\gamma_1})=\sigma^2_{\gamma_2}/(\sigma^2_{\alpha_2}+\sigma^2_{\beta_2}+\sigma^2_{\gamma_2})=\ldots=\sigma^2_{\gamma_M}/(\sigma^2_{\alpha_M}+\sigma^2_{\beta_M}+\sigma^2_{\gamma_M})\\ \sigma^2_{\alpha_1}/(\sigma^2_{\alpha_1}+\sigma^2_{\beta_1})=\sigma^2_{\alpha_2}/(\sigma^2_{\alpha_2}+\sigma^2_{\beta_2})=\ldots=\sigma^2_{\alpha_M}/(\sigma^2_{\alpha_M}+\sigma^2_{\beta_M})\\ \sigma^2_{\beta_1}/(\sigma^2_{\alpha_1}+\sigma^2_{\beta_1})=\sigma^2_{\beta_2}/(\sigma^2_{\alpha_2}+\sigma^2_{\beta_2})=\ldots=\sigma^2_{\beta_M}/(\sigma^2_{\alpha_M}+\sigma^2_{\beta_M}).\end{gathered}\tag{10}$$

Note that the test of $ICC_{\alpha p}$-equality is equivalent to the test of $ICC_{\beta p}$-equality.

The direction and *degree* of generalized and dyadic reciprocity are accurately depicted by correlations, not covariances. We, therefore, suggest to imply the Lagrange-multiplier method to the comparison of SRM-reciprocity coefficients as well. This yields

$$\sigma_{\alpha\beta_1}/\sqrt{\sigma^2_{\alpha_1}\sigma^2_{\beta_1}}=\sigma_{\alpha\beta_2}/\sqrt{\sigma^2_{\alpha_2}\sigma^2_{\beta_2}}=\ldots=\sigma_{\alpha\beta_M}/\sqrt{\sigma^2_{\alpha_M}\sigma^2_{\beta_M}}\tag{11}$$

for the restriction of the actor-partner correlation across a-groups to be equal and

$$\sigma_{\gamma\gamma_1}/\sigma^2_{\gamma_1}=\sigma_{\gamma\gamma_2}/\sigma^2_{\gamma_2}=\ldots=\sigma_{\gamma\gamma_M}/\sigma^2_{\gamma_M}\tag{12}$$

for the restriction of the relationship correlation across a-groups to be equal.The Likelihood Ratio (see Equation (9)) for a H_0-model containing any number c of these constraints within the Lagrange-multiplier and a comparison H_1-model not including the constraint(s) can be computed for a model comparison. While both models would contain an equal number of parameters estimated, the H_0-model would contain c degrees of freedom more due to the additional constraint(s), implying that the resulting test statistic will asymptotically follow a χ^2-distribution with c degrees of freedom under the H_0. Testing equality of a specific ICC or a specific correlation across a-groups implies $c = 1$ (see above).

4. Empirical Example

The data from Albright et al. (1988) on the interpersonal perception of intellect upon zero acquaintance was used for the present demonstration (see above). Perceived intellect was measured on a discrete 7-point-scale from 1 to 7 with higher numbers reflecting more of the trait. $N = 36$ rr-groups were investigated of which two contained six members, 21 contained five members, and 13 contained four members. Twenty-five rr-groups belonged to the first college and the remaining eleven belonged to the second college. We used all the described methods to test all SRM-parameters for differences between the two colleges. In the supplemental material, we present the data and an R (R Core Team, 2018) script for the analysis that was executed with OpenMx, version 2.19.5 (Neale et al., 2016). The estimator was FIML.

It should be noted that any ML-estimator for SEMs is conditioned on (multivariate) normality of the residuals (in the case of the SRM, these are the relationship effects; see Equation (1)). In an actual research scenario, applicants need to be cautious when discrete manifest variables—as in this example—enter or are planned to enter such an analysis. Discrete scales, by definition, cannot follow (continuous) normal distributions and strong deviations of the residuals from approximate normal distributions will result in a biased likelihood and underestimation of parameters and standard errors. However, discrete ordinal manifest variables may be used, as long as at least five response categories are given and distributions of residuals are approximately normal (which we assume for the data of this illustration; see the bottom of the R script from the supplemental material for an illustration on assessing (non-)normality of SRM-residuals with the current example). For a discussion on this topic, we refer to Finney and DiStefano (2006).

4.1. Results

Table 2 displays the estimates of all SRM-parameters for the two colleges and the parameter comparisons. The grand mean was rather high (>5) in both colleges, implying that students generally perceived each other as quite intellectual. The rr-group variance was significantly larger than zero only in the first college ($\sigma^2_{\mu_1} = 0.094$, $p = .044$). Note that the rr-groups from the first college distribute across three different psychology courses, whereas only one course was investigated in the second college (see the codebook for the data at: http://davidakenny.net/srm/srmdata.htm). Further investigations could clarify whether rr-group-level differences of average ratings in the first college are grounded in the different courses. Actor-effect and relationship-effect variances were substantial for both colleges, whereas the partner-effect variances were not significantly different from zero. Uniqueness accounted for more than half of the dyadic variance in both colleges. Assimilation accounted for around 90% of the person-level variance in both colleges. None of the SRM-covariance parameters were significantly different from zero in any college.

Table 2. Results of the mgSEM analysis of the data from Albright et al. (1988).

SRM-parameter	Estimate (college 1)	Estimate (college 2)	Abs. difference	W	Estimate (equality)	χ^2 ($df = 1$)
$\bar{\mu}$	5.251	5.264	0.013	0.077 (p = .469)	5.254	0.006 (p = .939)
σ^2_μ	0.094*	0.030	0.064	0.496 (p = .310)	0.085*	0.214 (p = .644)
σ^2_α	0.263**	0.628**	0.364	1.860 (p = .031)	0.353**	5.103 (p = .024)
σ^2_β	0.029	0.065	0.036	0.523 (p = .301)	0.037	0.294 (p = .588)
σ^2_γ	0.507**	0.759**	0.252	2.234 (p = .013)	0.591**	6.263 (p = .012)
ICC_α	.329	.432			.365	1.072 (p = .301)
ICC_β	.036	.045			.035	0.002 (p = .962)
ICC_γ	.634	.523			.596	0.967 (p = .325)
$ICC_{\beta p}$	.099	.094			.096	0.002 (p = .962)
$\sigma_{\alpha\beta}$	−0.024 (−.273)	0.096 (.477)			.058	2.053 (p = .152)
$\sigma_{\gamma\gamma}$	−0.055 (−.109)	−0.129 (−.170)			−.126	0.142 (p = .706)

Notes. Abs. difference: absolute difference (only depicted for estimates involved in the Wald-test); *W*: Wald statistic (*z*-value); Estimate (equality): estimate for the parameter in the H_0-model in which its value is restricted to equality across a-groups. χ^2 depicts the likelihood ratio between the H_1 and H_0 model. For ICC_α – $\sigma_{\gamma\gamma}$ the Lagrange-multiplier was used for the restrictions. Numbers in parentheses next to covariance parameters depict correlations. Only values for the absolute variances and the covariances were tested against 0 with: $^*p < .05$; $^{**}p < .001$.

The only significantly different parameter estimates were found for the actor-effect variance [$\Delta_{\text{college 2 – college 1}} = 0.364$; $W = 1.86$, $p = .031$, $\chi^2(1) = 5.103$, $p = .024$] and the relationship-effect variance [$\Delta_{\text{college 2 – college 1}} = 0.252$; $W = 2.234$, $p = .013$, $\chi^2(1) = 6.263$, $p = .012$]. It follows, that inter-individual differences with regards to perceiving others as intellectual and inter-relational differences with regard to perceiving a specific person as intellectual were more pronounced in the second college. Nonetheless, the relative degree to which individuals assimilated their perception across different persons as contrasted to consensus and unique dyadic perceptions was not significantly higher in the second college (ICC_α); neither was the relative degree to which this perception depended on dyadic relationships as contrasted to person-specific characteristics (ICC_γ).

5. Simulation Study

The aim of the simulation study was to determine Type 1 error-rate recoverability and statistical power for multiple group comparisons of the main SRM-parameters under varying conditions of rr-group sizes and number of rr-groups. We set $\bar{\mu}$ and σ^2_μ to zero in all populations and fixed them to zero in all estimations, since—based on our reasoning outlined above—we deem the remaining SRM-parameters as the actually interesting ones when comparing group dynamics. For the analysis of Type I error rates, a base population model was defined to hold in two a-groups. The population parameters are presented in the second column of Table 3. For simplicity, actor-, partner, and relationship-effect variances were selected to sum up to one. The exact values were chosen based on previous simulation studies of the SRM (Nestler, 2016, 2018; Nestler et al., 2020) and the well-replicated finding that dyadic interactions often are mostly characterized by uniqueness, followed by assimilation and, lastly, consensus (for an overview see Kenny, 2020). The actor-partner covariance and relationship covariance were chosen so that the respective correlations equal .2 and .1. For the comparison of absolute variances and correlations, two different population models were created for the second a-group, one for a "medium effect" condition and one for a "large effect" condition. The population parameters for these models are depicted in the third and fourth columns of Table 3, respectively. Absolute variance parameter values were 1.5 times larger in the medium effect condition and 2 times larger in the large effect condition compared to the base condition. These values were based on the median and average variance ratio, respectively, Keselman et al. (1998) found across empirical studies that employed analyses of variance for experimental groups. The differences of the correlations were chosen to reflect medium and large absolute values for correlations (Gignac & Szodorai, 2016, see also Cohen, 1988).

For the comparisons of ICCs, there were also a "medium" and a "large"-condition, but a different population model was defined for every investigated ICC (see Table 3). An ICC was increased by decreasing the respective other absolute variance parameters. The ratio of the decreased parameters was maintained. A "medium" effect was an increase of the ICC by .1 and a "large" effect was an increase of .2. The covariance parameters were changed so that the correlations of .2 and .1 for the actor-partner covariance and the relationship covariance, respectively, were maintained from the base model.

We used two different conditions of rr-group sizes: 5 and 13 (see also Kenny et al., 2006, p. 215). Four different conditions of rr-groups per a-group were analyzed: 1 *vs.* 5, 5 *vs.* 5, 10 *vs.* 10, and 20 *vs.* 20. These values were chosen based on practical restrictions SRM-researchers are likely to face in the conduction of their research.

The simulation study yields a 4 (number of rr-groups) × 2 (members per rr-group) × 3 (effect size: "none," "medium," "large") design. For the comparisons of absolute variances, the design is augmented by the additional two-level factor "test-procedure" (Wald *vs.* LR).

For every condition, 1,300 samples were drawn from the population. All solutions that yielded a warning message (e.g., non-positive definite covariance matrices of latent variables) were defined as invalid and discarded. The first 1,000 valid results were analyzed. The surplus of 300 samples per condition was drawn to account for the potential problem of low solution propriety with <1,000 valid results. However, this problem was not encountered as all solution propriety rates were >.90.

Table 3. Values for SRM-variance and covariance population parameters in the simulation study.

				rv-medium				rv-large			
par	base	av-medium	av-large	$ICC_{\beta p}$	ICC_{α}	ICC_{β}	ICC_{γ}	$ICC_{\beta p}$	ICC_{α}	ICC_{β}	ICC_{γ}
σ^2_α	.350	.525	.700	.225		.185	.233	.150		.115	.150
σ^2_β	.150	.225	.300		.099		.1		.066		.064
σ^2_γ	.500	.750	1		.329	.265			.220	.164	
$\sigma_{\alpha\beta}$	.046 (.200)	.137 (.400)	.275 (.600)	.037	.037	.033	.031	.030	.030	.026	.020
$\sigma_{\gamma\gamma}$	.050 (.100)	.225 (.300)	.500 (.500)	.05	.033	.026	.05	.050	.022	.016	.050

Notes. par: parameter; base: base-model [equal in both a-groups]; av: model for the comparison of absolute variances; rv: model for the comparison of relative variances; medium: model for the second a-group with medium effects; large: model for the second a-group with large effects. Empty cells hold the same value as the base-model. Numbers in parantheses depict correlations. In the rv-models, the correlations of $r_{\alpha\beta} = .200$ and $r_{\gamma\gamma} = .100$ from the base-model are maintained. Values are rounded to the third decimal.

Table 4. Results of the simulation study for the detection of group differences in absolute values of SRM variance parameters.

			Number of rr-groups			
Parameter	Members per rr-group	Test	1 *vs.* 5	5 *vs.* 5	10 *vs.* 10	20 *vs.* 20
σ^2_α	5	W	.897/.160/.181	.969/.071/.126	.965/.110/.290	.955/.278/.578
		LR	.955/.046/.073	.967/.062/.129	.971/.101/.282	.970/.194/.557
	13	W	.943/.269/.461	.943/.255/.643	.957/.506/**.927**	.953/**.821**/**.999**
		LR	.967/.070/.189	.971/.200/.587	.986/.425/**.890**	.964/.750/**.993**
σ^2_β	5	W	.941/.087/.100	.967/.034/.057	.964/.064/.119	.964/.129/.341
		LR	.966/.042/.041	.975/.042/.075	.978/.060/.132	.975/.086/.229
	13	W	.930/.192/.372	.971/.175/.491	.947/.381/**.821**	.952/.667/**.982**
		LR	.972/.057/.129	.972/.157/.446	.981/.316/.754	.976/.586/**.974**
σ^2_γ	5	W	.918/.246/.420	.967/.276/.551	.969/.504/**.910**	.948/.794/**.993**
		LR	.986/.034/.109	.974/.195/.523	.980/.410/**.861**	.971/.722/**.993**
	13	W	.957/**.879**/**.999**	.962/**.997**/**1**	.952/**1**/**1**	.953/**1**/**1**
		LR	.974/.734/**.997**	.979/**.999**/**1**	.975/**1**/**1**	.974/**1**/**1**

Notes. 1,000 samples were analyzed for every condition. W: Wald-test; LR: Likelihood-Ratio test. The first number in every cell depicts the relative number of correct maintenances of the null hypothesis (no differences between the two a-groups). The second number in every cell depicts the power (relative number of correct rejections of the null hypothesis) for medium effect sizes. The third number in every cell depicts the power for large effect sizes. Nomial Type I error = .025 (one-tailed). Power values >.80 are printed in bold.

The dependent variable for conditions with effect size "none" are rates of correct maintenances of the null hypotheses (no difference between the groups): 1—Type I error rate. For effect sizes conditions of "medium" and "large," the dependent variable was statistical power (correct rejection of the null hypothesis): 1—Type II-error rate. The nomial Type I error was set to .025 for a one-tailed test. The simulation was executed in R. The multivariate normally distributed populations were created with mvtnorm, version 1.1.1 (Genz et al., 2020), and the models were estimated with OpenMx, version 2.19.5 (Neale et al., 2016) using FIML. The R-scripts can be obtained from https://github.com/Jendryczko/mgSEM-SRM-simulation-study.

5.1. Results and Discussion

Table 4 displays the results for the absolute variance parameter comparisons. Type I error rates were mostly reproduced quite accurately. The Wald test yielded substantially more false rejections than the LR test but also had a higher power. However, in most cases when the power of the Wald test was sufficient, the power of the LR test was also acceptable (>.80), with the only exception being the comparison of the relationship effect-variance on 13 rr-group members and 1 *vs.* 5 a-groups with a medium effect size. It seems recommendable to always consider both, the Wald and the LR test and to see if they yield the same conclusion. At least 10 rr-groups in both a-groups are necessary to detect relationship-effect variance differences when every rr-group holds only five members. With 13 persons per rr-group, medium differences of the relationship-effect variance can be detected in almost all cases with two a-groups each containing only five rr-groups per a-group. If the difference is large, then two a-groups each containing 10 rr-groups with 13 members are sufficient to detect group differences for every variance parameter with acceptable power.

Table 5 presents the results for comparisons of ICCs and correlations. Type I error rates were acceptable. Medium effects were not sufficiently detected. With large rr-groups and effect sizes, a comparison of 5 *vs.* 5 rr-groups had sufficient power to detect differences in uniqueness and the relationship-correlations (for the latter parameter, even a 1 *vs.* 5 comparison suffices). 20 *vs.* 20 rr-groups containing 13 persons each were necessary to detect a larger difference of person-exclusive consensus. No condition came close to holding sufficient power for the detection of differences in the actor-partner correlation.

These results are in line with previous simulation studies on the SRM that showed that fewer rr-groups containing more members pose the statistically superior condition than a larger number of small rr-groups and that statistical properties for the estimation of dyad-level effects are generally better than for the estimation of person-level effects (Kenny et al., 2006; Lashley & Kenny, 1998; Nestler, 2016, 2018; Nestler et al., 2020). Note that our simulation conditions do not necessarily reflect "medium" or "large" differences of SRM-parameters that exist empirically. In a lot of cases, differences between a-

Table 5. Results of the simulation study for the detection of group differences in SRM ICCs and correlation parameters.

Parameter	Members per rr-group	Number of rr-groups 1 *vs.* 5	5 *vs.* 5	10 *vs.* 10	20 *vs.* 20
$ICC_{\beta p}$	5	.953/.052/.051	.970/.043/.075	.969/.048/.122	.974/.068/.257
	13	.975/.054/.134	.976/.099/.369	.975/.191/.687	.970/.371/**.914**
ICC_{α}	5	.946/.056/.103	.975/.060/.178	.972/.075/.350	.975/.184/.672
	13	.971/.080/.231	.974/.213/.721	.981/.385/**.954**	.970/.648/**.996**
ICC_{β}	5	.961/.068/.100	.971/.070/.236	.969/.122/.437	.979/.202/.751
	13	.977/.099/.354	.980/.299/**.874**	.972/.544/**.984**	.972/.763/**1**
ICC_{γ}	5	.976/.025/.059	.978/.055/.123	.967/.085/.288	.971/.150/.549
	13	.978/.070/.319	.977/.213/**.804**	.973/.460/**.982**	.973/.683/**1**
$r_{\alpha\beta}$	5	.934/.045/.064	.938/.058/.067	.946/.052/.068	.939/.038/.054
	13	.968/.050/.175	.961/.050/.138	.963/.063/.150	.969/.046/.163
$r_{\gamma\gamma}$	5	.970/.050/.114	.956/.054/.098	.953/.052/.107	.969/.045/.098
	13	.972/.214/**.847**	.983/.241/**.846**	.971/.215/**.828**	.977/.244/**.849**

Notes. 1,000 samples were analyzed for every condition.
$r_{\alpha\beta}$: actor-partner correlation; $r_{\gamma\gamma}$: relationship correlation. The first number in every cell depicts the relative number of correct maintenances of the null hypothesis (no differences between the two groups). The second number in every cell depicts the power (relative number of correct rejections of the null hypothesis) for medium effect sizes. The third number in every cell depicts the power for large effect sizes. Nomial Type I error = .025 (one-tailed). Power values >.80 are printed in bold.

groups might be much larger and easier to detect. Meta-analyses could establish often found differences of population parameters of the SRM that can be used to investigate necessary sample sizes for sufficient power. This will be crucial, especially for the actor-partner correlation, as differences in this parameter seem to be the most challenging to detect.

6. General Discussion

The present contribution illustrated various methods for the comparison of SRM parameters across different groups within a mgSEM framework. For the comparison of absolute variances, the Wald-test can be applied. However, this test seems to produce higher Type I error rates than nomial. The likelihood-ratio test poses the safer alternative that can also be employed to comparisons of SRM-correlation coefficients and ICCs when Lagrange-multipliers are used to impose the parameter restrictions between the a-groups. Our simulation studies further showed that sufficient power depends mostly on a sufficient number of rr-group members.

The SRM-effects were multivariate normally distributed in our simulated populations and the standard LR test might be insufficient when this is not the case. Alternatives can be probed in future studies. Scaling-corrected likelihood ratio tests in the framework of robust maximum likelihood estimation (MLR) are standard procedures for such cases (Satorra & Bentler, 2001). One might also consider bootstrapping as a non-parametric approach.

With our group-therapy example (see above) we already alluded to the potential of the approach to be extended to structural equation modeling of causal treatment effects (e.g., Mayer, 2019; Steyer, 2005) on dyadic interactions. Adding several points in time to the model is straightforward, as a longitudinal SEM SRM with a saturated latent variable covariance structure is statistically identical to the multivariate SRM (Nestler, 2018). The general SEM framework is also flexible enough for the inclusion of growth curves (Nestler et al., 2017) and to model covariates of SRM-effects (see also Lüdtke, et al., 2018) that need to be accounted for to increase the validity of causal inferences of the treatment.

Lastly, we would like to emphasize that the current contribution focused on the standard single-variable SRM, but that the proposed ideas also apply to more sophisticated models aiming at measuring actor-, partner-, and relationship-effects with several indicators (e.g., Bonito & Kenny, 2010; Nestler et al., 2020). In this context, a self-evident application of the multiple group comparison method is the investigation of measurement invariance across groups (Meredith, 1993). However, such models require very large sample sizes. Moreover, in the testing of measurement invariance, the null hypothesis constitutes the aimed for hypothesis. Even larger sample sizes are required to achieve sufficient power for falsification. The conduction of such large-scale studies will pose a challenge.

Acknowledgments

We would like to thank Johanna Eiben for her help during the literature review for this article.

Disclosure Statement

We have no conflict of interest to disclose. The research for this article was not funded by external sources. Figure 1 was reproduced and altered from a different manuscript written by the author that is currently under review. The Open Access data used in this article for illustration purposes was retrieved from http://davidakenny.net/srm/srmdata.htm (data-name: "Zero") and written permission to use it was inquired from David A. Kenny via email. The data is provided in the supplemental material to this article.

ORCID

David Jendryczko http://orcid.org/0000-0002-5984-8765

References

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, *55*, 387–395.

Betts, L. R., Rotenberg, K. J., Trueman, M., & Stiller, J. (2012). Examining the components of children's peer liking as antecedents of school adjustment. *The British Journal of Developmental Psychology*, *30*, 303–325.

Bonito, J. A., & Kenny, D. A. (2010). The measurement of reliability of social relations components from round-robin designs. *Personal Relationships*, *17*, 235–251.

Buist, K. L., Reitz, E., & Deković, M. (2008). Attachment stability and change during adolescence: A longitudinal application of the social relations model. *Journal of Social and Personal Relationships*, *25*, 429–444.

Christensen, P. N., & Feeney, M. E. (2016). Using the social relations model to understand dyadic perceptions within group therapy. *Group Dynamics: Theory, Research, and Practice*, *20*, 196–208.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*, 880–896.

Eichelsheim, V. I., Buist, K. L., Deković, M., Cook, W. L., Manders, W., Branje, S. J. T., … Meeus, W. H. J. (2011). Negativity in problematic and nonproblematic families: A multigroup social relations model analysis with structured means. *Journal of Family Psychology*, *25*, 152–156.

Ervin, J., & Bonito, J. A. (2014). A review and critique of partner effect research in small groups. *Small Group Research*, *45*, 603–632.

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Information Age Publishing.

Flores, P. J. (2001). Addiction as an attachment disorder: Implications for group therapy. *International Journal of Group Psychotherapy*, *51*, 63–81.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2020). *mvtnorm: Multivariate normal and t distributions*. R package version 1.1-1. http://CRAN.R-project.org/package=mvtnorm

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78.

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

Kenny, D. A., Gomes, S. B., & Kowal, C. (2015). The intergroup social relations model: ISRM. *Group Dynamics: Theory, Research, and Practice*, *19*, 152–165.

Kenny, D. A., & Livi, S. (2009). A componential analysis of leadership using the social relations model. *Multi-Level Issues in Organizational Behavior and Leadership*, 147–191.

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

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., … Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, *68*, 350–386.

Lam, C. K., Van der Vegt, G. S., Walter, F., & Huang, X. (2011). Harming high performers: A social comparison perspective on interpersonal harming in work teams. *The Journal of Applied Psychology*, *96*, 588–601.

Lashley, B. R., & Kenny, D. A. (1998). Power estimation in social relations analyses. *Psychological Methods*, *3*, 328–338.

Loeys, T., Fonteyn, M., & Loncke, J. (2021). A practical tool for family assessment based on the social relations model. *Frontiers in Psychology*, *12*, 699831.

Lüdtke, O., Robitzsch, A., & Trautwein, U. (2018). Integrating covariates into social relations models: A plausible values approach for handling measurement error in perceiver and target effects. *Multivariate Behavioral Research*, *53*, 102–124.

Marcus, D. K. (1998). Studying group dynamics with the social relations model. *Group Dynamics: Theory, Research, and Practice*, *2*, 230–240.

Marcus, D. K. (2021). Studying group dynamics with the social relations model revisited. *Group Dynamics: Theory, Research, and Practice*, *25*, 250–257.

Markin, R. D., & Kivlighan, D. M. Jr. (2008). Central relationship themes in group psychotherapy: A social relations model analysis of transference. *Group Dynamics: Theory, Research, and Practice*, *12*, 290–306.

Mayer, A. (2019). Causal effects based on latent variable models. *Methodology*, *15*, 15–28.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., … Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*, 535–549.

Nestler, S. (2016). Restricted maximum likelihood estimation for parameters of the social relations model. *Psychometrika*, *81*, 1098–1117.

Nestler, S. (2018). Likelihood estimation of the multivariate social relations model. *Journal of Educational and Behavioral Statistics*, *43*, 387–406.

Nestler, S., Geukes, K., Hutteman, R., & Back, M. D. (2017). Tackling longitudinal round-robin data: A social relations growth model. *Psychometrika*, *82*, 1162–1181.

Nestler, S., Lüdtke, O., & Robitzsch, A. (2020). Maximum likelihood estimation of a social relations structural equation model. *Psychometrika*, *85*, 870–889.

Olsen, J. A., & Kenny, D. A. (2006). Structural equation modeling with interchangeable dyads. *Psychological Methods*, *11*, 127–141.

Oosterhof, A., Van der Vegt, G. S., Van de Vliert, E., & Sanders, K. (2009). Valuing skill differences: Perceived skill complementarity and dyadic helping behavior in teams. *Group & Organization Management*, *34*, 536–562.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

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.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507–514.

Sfetcu, R. (2013). Perceiver and target effects evaluation the professional competencies of others. Does it have an impact on inter-profession al collaboration? *Human Resource Management: Management in Health*, *17*, 19–24.

Stas, L., Schönbrodt, F., & Loeys, T. (2015). Getting the most out of family data with the R package fSRM. *Journal of Family Psychology*, *29*, 263–275.

Steyer, R. (2005). Analyzing individual and average causal effects via structural equation models. *Methodology*, *1*, 39–54.

Voelkle, M. C., Oud, J. H. L., von Oertzen, T., & Lindenberger, U. (2012). Maximum likelihood dynamic factor modeling for arbitrary N and T using SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*, 329–350.

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.

Yuan, K.-H., & Liu, F. (2021). Which method is more reliable in performing model modification: Lasso regularization or Lagrange multiplier test? *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 69–81.