METHODOLOGY
European Journal of Research Methods for the Behavioral and Social Sciences

PsychOpen GOLD

Original Article

Scaling Metric Measurement Invariance Models

Eric Klopp [1], Stefan Klößner [2]

[1] *Department of Education, Saarland University, Saarbrücken, Germany.* **[2]** *Department of Economics, Saarland University, Saarbrücken, Germany.*

Methodology, 2023, Vol. 19(3), 192–227, https://doi.org/10.5964/meth.10177

Received: 2022-08-30 • **Accepted:** 2023-07-03 • **Published (VoR):** 2023-09-29

Handling Editor: Jost Reinecke, Bielefeld University, Bielefeld, Germany

Corresponding Author: Eric Klopp, Department of Education, Bldg. A 4.2, Saarland University, 66123 Saarbrücken, Germany. E-mail: e.klopp@mx.uni-saarland.de

Supplementary Materials: Materials [see Index of Supplementary Materials]

Abstract

This paper aims at clarifying the questions regarding the effects of the scaling method on the discrepancy function of the metric measurement invariance model. We provide examples and a formal account showing that neither the choice of the scaling method in general nor the choice of a particular referent indicator affects the value of the discrepancy function. Thus, the test statistic is not affected by the scaling method, either. The results rely on an appropriate application of the scaling restrictions, which can be phrased as a simple rule: "Apply the scaling restriction in one group only!" We develop formulas to calculate the degrees of freedom of χ^2-difference tests comparing metric models to the corresponding configural model. Our findings show that it is impossible to test the invariance of the estimated loading of exactly one indicator, because metric MI models aimed at doing so are actually equivalent to the configural model.

Keywords

metric measurement invariance, partial measurement invariance, scaling methods, referent indicator, equivalence

Measurement invariance analysis has become a widely used tool in many areas of applied research, e.g., psychology, sociology, and organizational research, and has received a lot of attention in methodological research, e.g., Davidov et al. (2018), Greiff and Scherer (2018), Millsap (2011), and van de Schoot et al. (2015); see Leitgöb et al. (2023) for a recent overview of measurement invariance analysis in the social sciences. The goal of measurement invariance (MI) analysis is to investigate whether measurement properties

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0, which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

of a latent variable's indicators are the same across two or more groups. MI analysis has many facets, e.g., which method is the best when looking for invariant indicators. However, each of these methods requires the application of a certain scaling method. Without scaling, the respective models could not be uniquely estimated. The focus of this paper is on this specific requirement of each MI investigation, i.e. how to set scaling restrictions appropriately when investigating the invariance of factor loadings.

There are several levels of MI that are investigated consecutively (e.g., Brown, 2015, pp. 242–243). The lowest level is configural MI, where the factor structure is the same in all groups. At the next level, metric MI, the invariance condition states additionally that the loadings are equal across groups. The third level is scalar MI, which means that the intercepts are invariant across groups[1] . In this paper, we focus on the metric MI model and elaborate on the question of how to scale this model. In general, we examine whether the scaling method affects the discrepancy function, and in turn the resulting test statistic of a model. In particular, we examine whether the selection of a particular referent indicator or, more generally, the selection of certain scaling method affects the discrepancy function, and in turn the resulting test statistic, for the metric MI model.

The test statistics of both the configural and metric MI model are assumed to be χ^2-distributed with df degrees of freedom under the corresponding null hypothesis. The test statistic's numeric value results from the estimation process, in which a discrepancy function F is minimized. According to the procedure in Vandenberg and Lance (2000, p. 56), testing for metric MI requires the following steps: In the first step, the configural model is tested. In the second step, the metric model is tested. If the test statistics indicate in both steps that the model fits the data, then a χ^2-difference test is conducted in the third step. This last test evaluates whether the metric MI conditions hold. Thus, MI analysis draws on testing single models as well as on testing nested models.

In addition to the level of MI analysis, there is also a distinction between *full* and *partial* metric MI models. In the *full* MI model, all loadings are assumed to be invariant across groups, while in *partial* MI models, one or more loadings are free to vary between the groups.

As a metric MI model entails latent variables, it is a well known fact that scaling restrictions must be placed, otherwise the model would not be identified. In this paper, we use the term scaling *restriction* to refer to the fixation of a certain parameter to a given value imposed by the application of a certain scaling method and use the term invariance *condition* to refer to the imposed equality of the loadings in a metric MI model. This distinction follows Wu and Estabrook (2016) who introduced a similar distinction between scaling restriction and invariance conditions. Finally, we use the term *constraints* to refer to both, i.e., the combination of the invariance conditions and

1) Sometimes, the terms weak, strong, and strict MI are used instead of configural, metric, and scalar MI (e.g., Kline, 2016, p. 396).

the scaling restriction in a model. The Fixed Marker (FM) scaling (cf. Kline, 2016, pp. 199–200) is the most commonly applied scaling method and it is the de facto standard in most structural equation software. According to Raykov et al. (2012, p. 968), the FM method restricts one of the estimated loadings of each latent variable to 1 in each group. The respective manifest variable is then called the referent indicator (RI). From that description, it is obvious that the FM scaling method has as many different variants as there are indicators for a latent variable. For instance, when there are 6 indicators, then each of these indicators' estimated loadings can be restricted to 1. In this paper, if the first indicator is the RI, we use the abbreviation FM_1. If the second indicator is the RI, we write FM_2, and so on.[2] Usually, it is suggested that the indicator that best represents the latent variable should be the RI (e.g., Brown, 2015, pp. 242–243).

In the context of metric MI analysis, the use of the FM method has been critically discussed in textbooks.[3] According to Brown (2015, p. 271; cf., Cheung and Rensvold, 1999, pp. 8–9), there may be difficulties in multiple group settings regarding the selection of an RI. For instance, Brown (2015) mentions that non-invariance of the RI may not be detected in a metric MI setting, because it is restricted to 1 in every group, and as a consequence thereof, there is the implicit assumption that the RI is invariant across groups. Similarly, Kline (2016, p. 405) argues that the selection of an RI is arbitrary and that the fixation of the RI to 1 over the groups is tantamount to an assumption of invariance of the RI. Furthermore, Kline argues that the RI is excluded from the test of metric MI, but the RI choice should not affect the overall model fit most of the time. Raykov et al. (2020) provide an example in which a metric MI model is estimated using an non-invariant RI. In a Monte Carlo simulation, the metric MI model failed to reveal the non-invariance of the RI by means of the test statistic most of the time. However, the authors did not consider choosing an invariant RI.

Johnson et al. (2009, p. 654) found in a Monte Carlo simulation for the full metric MI model that when RIs are non-invariant, this non-invariance does not distort the results of metric MI tests. The authors state that differences between groups on the RIs were transferred to other indicators via the constraints on the RI to be equal to 1 in both groups, which in turn affected the estimated loadings of all other indicators by setting a scale. Furthermore, the authors summarize that, in general, the tests of the full metric MI model were accurate in detecting non-invariance when the RI differed across groups, regardless of whether or not other indicators also differed. Given these results, Johnson et al. (2009, p. 644) argue that all estimated loadings are rescaled relative to the RI, by a magnitude of $1/\lambda_{RI}$ (where λ_{RI} denotes the "true" loading of the RI) and although the RI is non-invariant across groups, the parameter estimates will be adjusted by different

2) In this paper, we assume that for all latent variables in a model, the same indicator is used as a RI.

3) In the following, we concentrate the citations mostly on established and contemporary textbooks, because they reflect a synthesis of the current literature on the topic.

scaling constants and the loadings are expressed in two different metrics (Johnson et al., 2009, p. 644). This kind of rescaling is also mentioned by Raykov et al. (2012, Appendix A). These authors state that setting the RI to 1 yields a rescaling of the remaining loadings.

The concerns and findings mentioned above were in the context of the full metric MI model. For *partial* metric MI, Johnson et al. (2009) found simulation evidence that the choice of the RI affects the test results.[4] Vandenberg and Lance (2000, pp. 46–47) state that in partial metric MI settings, problems may arise when an RI is in reality non-invariant because that yields different metrics for a latent variable over the groups (cf., Brown, 2015).

Up to here, we have only considered the choice of the RI in the FM scaling method, but there are two other methods of scaling latent variables in multiple group settings with metric MI conditions (cf. Kline, 2016, pp. 199–200). These are the Effects Coding (EC) method and the Reference Group (RG) method.

The EC method (Little et al., 2006) restricts the estimated loadings by enforcing that their sum equals the number of indicators. This restriction is equivalent to the restriction that the average of the estimated loadings of a latent variable equals 1.

The RG method (cf., Kline, 2016, p. 393, Lee et al., 2011, p. 60, Raykov et al., 2012, p. 961, Wu & Estabrook, 2016), which is the multiple group version of the Fixed Factor (FF) scaling, restricts the variances of the latent variables to 1 in the reference group only, while being freely estimated in the remaining groups. As noted by Raykov et al. (2012, p. 969), the identification of this model is achieved by the imposed equality constraints of the estimated loadings in the remaining groups. Put differently, the metric MI conditions *propagate* the scaling to the other groups. This is in contrast to the FM method, in which an RI is set to 1 in each group (see above).

Obviously, due to the lack of an RI, neither EC nor RG contain any similar invariance assumption. Additionally, the FM scaling method only provides estimated loadings for all indicators but not the RI, whereas the EC or RG provide estimated loadings for all indicators. Indeed, van de Schoot et al. (2012, p. 489) recommend using the RG method when the goal is to compare factor loadings across groups.

In a nutshell, what was said above provides an inconsistent picture on the issues involved in scaling (partial) metric MI models. These inconsistencies can be summed up in the following four questions:

1. Does it matter which indicator serves as RI or is the result of an MI analysis in sense of the model fit of the metric MI model independent of the RI choice?
2. More generally, does the choice of the scaling method in general affect the results of an MI analysis or do all scaling methods provide the same result?

4) Johnson et al. (2009) use the term ‘scale level’ to refer to a full metric MI setting, whereas they use the term ‘item-level’ for a partial metric MI model in which they intend to test the invariance of exactly one loading.

3. Do the same considerations apply in *full* and *partial* metric MI models or do we have to tell apart these settings?
4. Can the FM scaling method be applied by only restricting the RI in one group and propagating the scaling restriction via the MI condition like the RG method?

The fundamental topic of the first two questions is if there are any effects of the scaling method on the value of the discrepancy function F. All four questions are not only of interest for methodological considerations, but also for applied researchers who wish to conduct a MI analysis. This paper aims at clarifying these questions and thus aims to bring consistency into the literature concerning the choice of RIs in particular or the choice of a scaling method in general. The general approach of the paper is to present examples to empirically demonstrate some basic facts and ideas, and then to turn to formal considerations. However, the formal part is kept at the minimal required level.

In the following section, we provide the general setup and notations used in this paper. Then, we first present an example showing that neither the scaling method in general nor the choice of an RI in particular affects the discrepancy function of the full metric MI model. We also use the example to introduce the concept of change of scale, which we make use of in the following section to provide a formal proof of the equivalence of all scaling methods for the full metric MI model. Additionally, we derive some corollaries about the resulting test statistic and other fit indices, showing that these are also unaffected by the scaling methods. Furthermore, we also provide a corollary about the degrees of freedom for the χ^2-difference test between the configural and the full metric MI model, which we then extend to partial metric MI models. As a consequence, we find that a partial metric MI model with only one invariance condition per latent variable is equivalent to the configural model. Thus, it is impossible to test the invariance of only one indicator (Steenkamp & Baumgartner, 1998). In the following section, we provide a further example, in which we firstly empirically demonstrate these corollaries. Again, the example illustrates that neither the scaling method in general nor the choice of an RI in particular affect the discrepancy function of the partial metric MI model. Afterwards, we provide a formal account for the partial MI model. In particular, we explain why the (non-)invariance of an RI, or more generally the data generating process, does not affect tests of either full or partial metric MI. Afterward, in the conclusion we elaborate on the results obtained in the paper and provide the answers to the four questions. Supplementary materials containing additional information and R scripts for the examples are available in Klopp and Klößner (2022) and referred to in the appropriate places in the text.

Setup

In this section, we introduce the notation we use throughout this article. Basically, we consider measurement models or confirmatory factor analysis models in multiple group settings of the form

$$\boldsymbol{X}_g = \boldsymbol{\Lambda}_g \boldsymbol{\xi}_g + \boldsymbol{\delta}_g \,. \tag{1}$$

Let p and m be the number of manifest and latent variables, resp., and let $g \in 1, \ldots, G$ be an index of the group membership. Then $\boldsymbol{X}_g$ is a $p \times 1$ vector of manifest variables, $\boldsymbol{\Lambda}_g$ is a $p \times m$ loading matrix, $\boldsymbol{\xi}_g$ is a $m \times 1$ vector of latent variables, and $\boldsymbol{\delta}_g$ is a $p \times 1$ vector of measurement residuals. Additionally, $\lambda_{j,i,g}$ represents the loading of the j-th indicator on the i-th latent variable in group g. The other model parameters are indexed analogously.

Under the standard assumptions for confirmatory factor analysis models (e.g., Bollen, 1989, pp. 233–234), this measurement model yields the following model-implied $p \times p$ covariance matrix for $\boldsymbol{X}_g$:

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

where $\boldsymbol{\Phi}_g$ is the $m \times m$ covariance matrix of the latent variables $\boldsymbol{\xi}_g$ and $\boldsymbol{\Theta}_g$ is the $p \times p$ covariance matrix of the measurement residuals $\boldsymbol{\delta}_g$, which is assumed to be a diagonal matrix containing the manifest residual variances. We collect the parameter matrices in the parameter vector $\theta = \left(\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_G, \boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_G, \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_G\right)$. To simplify our considerations, we assume a fully saturated model for the mean structure.

As is common in statistics, we distinguish between population models and estimated models (cf., Klopp & Klößner, 2021, pp. 185–187, Figure 4). In particular, we refer to a model in the form of Equation (1) as the population model that contains the data generating process (dgp). If the relation

$$\boldsymbol{\Lambda}_{g_1} = \boldsymbol{\Lambda}_{g_2} \quad \forall g_1 \neq g_2 \in 1, \ldots, G \tag{3}$$

holds within a dgp, then this dgp fulfills full metric invariance. In contrast, a dgp fulfills *partial* metric MI if

$$\lambda_{j,i,g_1} = \lambda_{j,i,g_2} \tag{4}$$

for all groups $g_1 \neq g_2$ and for some combinations of indicator j and latent variable i for which the corresponding nonzero loadings are invariant across groups, while some loadings are allowed to be non-invariant. We call the dgp either a full or a partial metric MI dgp.

Following the above-mentioned distinction between population and estimated models, the estimated model is stipulated by a researcher. We assume that the estimated

model is also a confirmatory factor analysis model of the same structure. The estimated model implied-covariance matrix is

$$\widehat{\boldsymbol{\Sigma}}_g = \widehat{\boldsymbol{\Lambda}}_g \widehat{\boldsymbol{\Phi}}_g \widehat{\boldsymbol{\Lambda}}_g' + \widehat{\boldsymbol{\Theta}}_g \,. \tag{5}$$

The parameter vector $\widehat{\theta} = \left(\widehat{\boldsymbol{\Lambda}}_1, \ldots, \widehat{\boldsymbol{\Lambda}}_G, \widehat{\boldsymbol{\Phi}}_1, \ldots, \widehat{\boldsymbol{\Phi}}_G, \widehat{\boldsymbol{\Theta}}_1, \ldots, \widehat{\boldsymbol{\Theta}}_G\right)$ collects the model matrices of the estimated model.

The invariance condition for *full* metric MI on the estimated loadings is

$$\widehat{\boldsymbol{\Lambda}}_{g_1} = \widehat{\boldsymbol{\Lambda}}_{g_2} \quad \forall g_1 \neq g_2 \in 1, \ldots, G \,. \tag{6}$$

For *partial* metric MI, these invariance conditions are replaced by the less restrictive invariance conditions

$$\widehat{\lambda}_{j,i,g_1} = \widehat{\lambda}_{j,i,g_2} \tag{7}$$

for all groups $g_1 \neq g_2$ and all combinations of indicator j and latent variable i for which the corresponding estimated nonzero loadings are assumed to be invariant across groups.

In the following, we will always explicitly use either the term full metric MI model or partial metric MI model. If we want to make assertions that refer to both types of models, we use the term *metric MI model*. Moreover, we call a model such as in Equation (5) in combination with invariance conditions as in Equation (6) or as in Equation (7) the *unscaled* model, cf. Klopp and Klößner (2021, pp. 185–187), as no scaling restrictions have been added yet.

To actually estimate the model, we have to apply a scaling method. It is important to note that the restrictions of the various scaling methods are applied to the *estimated* model (Klopp & Klößner, 2021, p. 185), but not to the dgp. We collect the scaling methods in the set

$$\mathcal{S} = \{\mathrm{FM}_1, \ldots, \mathrm{FM}_p, \mathrm{EC}, \mathrm{RG}\} \,. \tag{8}$$

To refer to any of the elements of $\mathcal{S}$, we use the generic notation S. The estimated model-implied covariance matrix then is

$$\widehat{\boldsymbol{\Sigma}}_g^S = \widehat{\boldsymbol{\Lambda}}_g^S \widehat{\boldsymbol{\Phi}}_g^S \widehat{\boldsymbol{\Lambda}}_g^{S'} + \widehat{\boldsymbol{\Theta}}_g \,. \tag{9}$$

We call a model as in Equation (9) a *scaled* model. Please notice that the estimated manifest residual variances $\widehat{\boldsymbol{\Theta}}_g$ in the scaled model do not have a superscript indicating the scaling method. This is because in order to estimate residual variances, no scaling method needs to be applied and estimated residual variances do not depend on the

scaling method employed, i.e., the estimated values are identical for all scaling methods S (Klößner and Klopp, 2019, p. 148; Klopp and Klößner, 2021, p. 185).

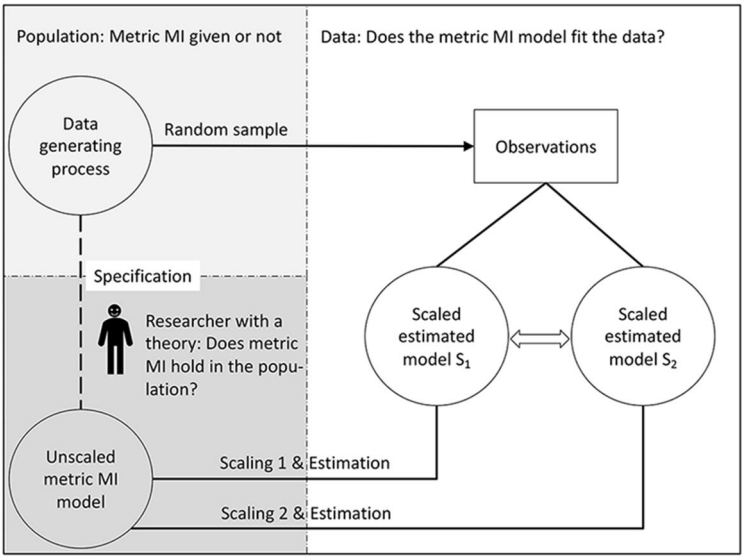
In the following, when estimating a metric MI model, we assume that configural invariance is given. Furthermore, we assume that the scaled model is estimated by minimizing a discrepancy function $F(\boldsymbol{S}_g, \widehat{\boldsymbol{\Sigma}}_g^S)$. The discrepancy function is based on a sample covariance matrix $\boldsymbol{S}_g$ and the model implied matrix $\widehat{\boldsymbol{\Sigma}}_g^S$. To keep the notation simple we write $F(\widehat{\theta})$ (cf., Bollen, 1989, p. 106, for the properties of a discrepancy function).

As the distinction between a dgp as well as unscaled and scaled estimated models is not common in the literature, we want to explain the relations between these models with the help of Figure 1 (cf., Klopp & Klößner, 2021, p. 186). Firstly, in the upper left corner, there is a *dgp* that functions in a population and from which a random sample of observations is drawn. The observations are collected in a data set that is later used in the estimation process, which is depicted on the right side of the figure. Secondly, in the lower left corner, there is a researcher who has a *substantial theory* about the metric MI in the population and is interested in answering the *substantial hypothesis* "Does metric MI hold in the population?" Importantly, the dgp is unknown to the researcher. To answer the question, the researcher specifies a statistical model like the one given in Eequation (5), with metric MI conditions of the form given in equation (6) or equation (7), i.e., the researcher specifies an *unscaled model*. This represents the researcher's *statistical hypothesis*. To actually estimate the model, the researcher has to apply scaling restrictions, i.e., the researcher creates a *scaled model*. This can be done with any one of the available scaling methods, of which the two possibilities S_1 and S_2 are depicted in the figure. With the scaled model, the statistical hypothesis of whether the metric MI model fits to the data is tested[5] and afterward, the researcher arrives at a conclusion about the substantial hypothesis. The figure illustrates two issues: Firstly, the model used by the researcher is a distinct entity from the dgp. Secondly, the researcher starts with invariance conditions in the unscaled model, to which the scaling methods are then added. Thus, the scaling of the model is not related to the dgp. We will now move on to our first example.

5) With this, we mean the sequence of tests as outlined in the introduction.

Figure 1

Relation Between the dgp, the Estimated Model and the Various Scaling Methods



Full Metric Measurement Invariance: An Example

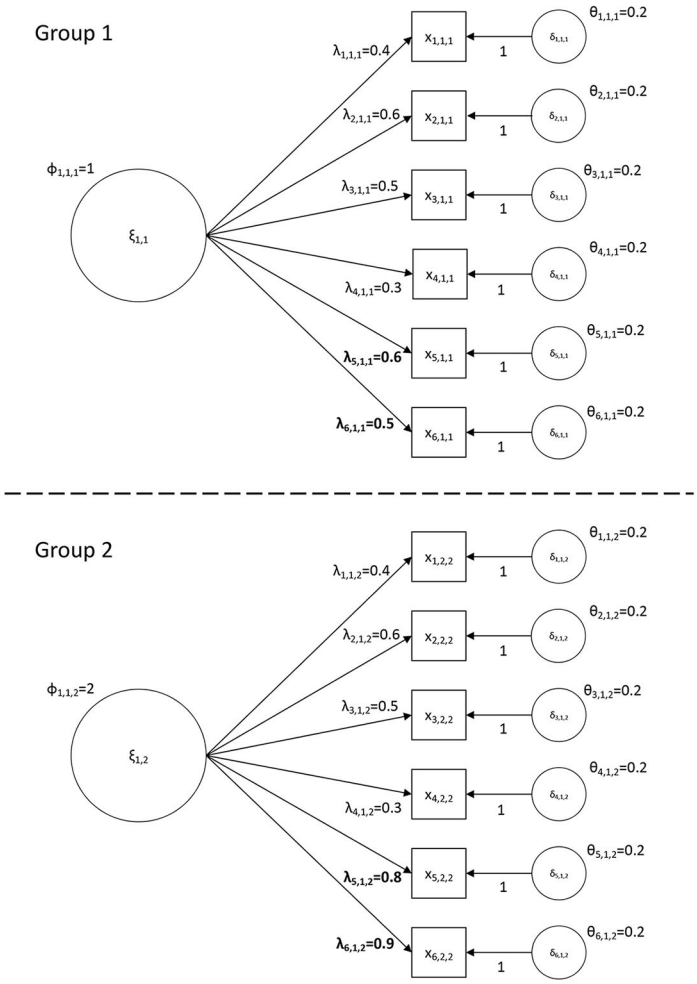
In the following example, we present a simple two-group case which demonstrates that regardless of the scaling method, the same value of the discrepancy function results when a full metric MI model is estimated. The aim of this section is to provide an intuitive and heuristic understanding, both of the equivalence of the scaling methods for the full metric MI model and of the concept of change of scale. A change of scale allows to convert the estimates obtained under a certain scaling method to those obtained under any other scaling method, without re-estimating the model. This concept is vital to the more formal account in the next section.

We consider a model with only one latent variable with six indicators in two groups, $g_0 = 1$ indicates the reference group and $g = 2$ indicates the focal group (cf., Holland & Thayer, 1988). The example's dgp is given by (see also Figure 2 for a graphical display of the dgp):

$$
\begin{aligned}
&\boldsymbol{\Lambda}_1 = (0.4, 0.6, 0.5, 0.3, 0.6, 0.5)', \\
&\boldsymbol{\Lambda}_2 = (0.4, 0.6, 0.5, 0.3, 0.8, 0.9)', \\
&\phi_{1,1,1} = 1.0, \\
&\phi_{1,1,2} = 2.0, \\
&\theta_{j,j,g} = 0.2,\ 1 \leq j \leq 6,\ \ g = 1, 2
\end{aligned}
\tag{10}
$$

Figure 2

CFA Model Depicting the dgp From Equation 10



Note. The loadings printed in bold are non-invariant in the dgp.

This is a *partial* metric MI dgp. The loadings of the last two indicators are non-invariant. Especially, both loadings differ to a different extent. If the concerns about non-invariant RIs in the FM scaling methods were justified, then the estimated values of the discrepancy function (on which the test statistic rests) would differ.

We now take the role of the researcher from Figure 1 and set up our unscaled model as in Equation (5) and with the full metric invariance condition $\widehat{\mathbf{\Lambda}}_{g_1} = \widehat{\mathbf{\Lambda}}_{g_2}$. In detail, this invariance condition entails the following set of invariance conditions:

$$\begin{aligned}
\widehat{\lambda}_{1,1,1} &= \widehat{\lambda}_{1,1,2} \\
\widehat{\lambda}_{2,1,1} &= \widehat{\lambda}_{2,1,2} \\
\widehat{\lambda}_{3,1,1} &= \widehat{\lambda}_{3,1,2} \\
\widehat{\lambda}_{4,1,1} &= \widehat{\lambda}_{4,1,2} \\
\widehat{\lambda}_{5,1,1} &= \widehat{\lambda}_{5,1,2} \\
\widehat{\lambda}_{6,1,1} &= \widehat{\lambda}_{6,1,2}
\end{aligned} \tag{11}$$

To scale the model, we apply each scaling method from the set $\mathcal{S}$ and start with the FM scaling method, using the first indicator as RI[6] , i.e., FM_1. The FM_1 scaling restriction is $\widehat{\lambda}_{1,1,1} = 1$. Considering the other indicators as RI, we can express all variants of the FM scaling method as

$$\widehat{\lambda}_{j,1,1} = 1 \quad \text{for} \quad 1 \leq j \leq 6. \tag{12}$$

Notably, the invariance condition in Equations (11) yields $\widehat{\lambda}_{j,1,2} = 1$ and we see that the invariance condition propagates the FM scaling restriction from the reference to the focal group.

For the EC scaling method, the scaling restriction we set in the reference group is

$$\widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 6. \tag{13}$$

As stated in the introduction, this scaling restriction is equivalent to the restriction that the average of the estimated loadings equals 1, i.e.,

$$\text{A}(\widehat{\lambda}_{\bullet,1,1}) := \frac{\widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1}}{6} = 1. \tag{14}$$

6) We only consider the case when the scaling restrictions are set in the reference group. The case when the scaling restrictions are set in the focal group, which means a role swap of both groups, can be easily implemented in the R code.

Again, the invariance conditions cause the propagation of the scaling restriction to the focal group.

The most important observation from these scaling examples is that the scaling restrictions are applied in one group only. As we have seen, a common feature of all scaling methods is the propagation of the scaling restrictions into the focal group via the invariance conditions. Because of its importance, we want to set up this rule in colloquial terms, which is *Apply the scaling restriction in one group only, the invariance conditions do the rest!*

Finally, for the RG scaling method, the scaling restriction in the reference group is

$$\widehat{\phi}_{1,1,1} = 1. \tag{15}$$

In contrast to the previous scaling methods, this restriction does not directly propagate to the focal group. However, the restriction on the estimated variance in the reference group scales the estimated loadings in this group indirectly. Due to the invariance conditions, this restriction is then propagated to the focal group, thus overall acting as a scaling restriction.

We now estimate the full metric MI model for ten different generated samples, using all scaling methods, i.e., the six different variants of the FM scaling as well as the EC and RG scaling method. We use the ML estimator. The sample size per group is $N = 300$, thus $N = 600$ in total. We chose this sample size because it reflects a number often achieved in real studies. All calculations for the data generation and model estimation were done in R (R Core Team, 2022), using the packages MASS (Venables and Ripley, 2002) and lavaan (Rosseel, 2012).

Regarding the degrees of freedom, there are $p(p+1)/2$ unique entries in the covariance matrix in each group, i.e., with $p = 6$ manifest variables there are 21 unique entries in each group and thus 42 unique pieces of information in total. In each group, we have to estimate 6 loadings, 6 manifest residual variances and 1 latent variance, i.e., 13 parameters per group and 26 parameters in total. Because we have a full metric MI model, there are 6 invariance conditions (see Equation 11). Additionally, there is 1 scaling restriction, regardless of the scaling method; see Equations (12), (13), and (15). Thus, there are 7 restrictions in total, which gives $26 - 7 = 19$ free parameters. This results in $df = 42 - 19 = 23$ degrees of freedom.

The results are presented in Tables 1 and 2. Table 1 shows the values of the ML discrepancy function resulting from the estimation, multiplied by the factor 10^6, while Table 2 shows the estimated loadings and latent variances for the first generated sample. In each sample, the value of the ML discrepancy function is the same regardless of the scaling methods. In summary, we can state as a first observation that neither the specific choice of an RI nor in general the choice of the scaling method affect the full metric MI model's resulting value of the discrepancy function. Consequently, the test statistic and all fit indices derived from the test statistic are also identical.

This result is in line with Johnson et al. (2009) and demonstrates empirically that any concerns about truly non-invariant RIs are unwarranted, at least in this example. Considering the distinction between population and specified models explains this finding. The non-invariance is in the dgp, not in the specified, scaled model, in which an invariance condition is set by the researcher. This restriction represents an invariance assumption that applies to the estimates only, it is used to test if this restriction mirrors the invariance properties of the dgp.

Table 1

Values of the Discrepancy Function Multiplied by the Factor 10^6 *for Each Scaling Method in the Full Metric MI Example*

Sample	FM_1	FM_2	FM_3	FM_4	FM_5	FM_6	EC	RG
1	58978	58978	58978	58978	58978	58978	58978	58978
2	94356	94356	94356	94356	94356	94356	94356	94356
3	62417	62417	62417	62417	62417	62417	62417	62417
4	72264	72264	72264	72264	72264	72264	72264	72264
5	66726	66726	66726	66726	66726	66726	66726	66726
6	85389	85389	85389	85389	85389	85389	85389	85389
7	58682	58682	58682	58682	58682	58682	58682	58682
8	59634	59634	59634	59634	59634	59634	59634	59634
9	116513	116513	116513	116513	116513	116513	116513	116513
10	75311	75311	75311	75311	75311	75311	75311	75311

Table 2

Estimated Loadings and Latent Variances for the First Sample in the Full Metric MI Example; $g = 1, 2$, $S \in \{FM_1, \ldots, FM_6, EC, RG\}$.

Loadings	FM_1	FM_2	FM_3	FM_4	FM_5	FM_6	EC	RG
$\widehat{\lambda}^S_{1,1,g}$	1.000	0.688	0.829	1.329	0.550	0.534	0.740	0.323
$\widehat{\lambda}^S_{2,1,g}$	1.453	1.000	1.204	1.931	0.799	0.775	1.076	0.470
$\widehat{\lambda}^S_{3,1,g}$	1.207	0.831	1.000	1.604	0.664	0.644	0.893	0.390
$\widehat{\lambda}^S_{4,1,g}$	0.752	0.518	0.623	1.000	0.414	0.401	0.557	0.243
$\widehat{\lambda}^S_{5,1,g}$	1.818	1.251	1.506	2.417	1.000	0.970	1.346	0.588
$\widehat{\lambda}^S_{6,1,g}$	1.874	1.290	1.553	2.492	1.031	1.000	1.388	0.606
$\widehat{\phi}^S_{1,1,1}$	0.104	0.220	0.152	0.059	0.345	0.367	0.191	1.000
$\widehat{\phi}^S_{1,1,2}$	0.345	0.728	0.502	0.195	1.140	1.211	0.629	3.302

Now, we turn to a second observation: Parameters estimated under one scaling method can be converted to those estimated under another scaling method (Klopp and Klößner, 2021, Proposition 2; Newsom, 2015, p. 4). We illustrate this conversion and convert the estimated parameters under FM_1 scaling to those obtained under RG and EC scaling for the first out of the ten generated samples. The estimated parameters under FM_1 scaling are:

$$\begin{aligned}
\widehat{\boldsymbol{\Lambda}}_g^{\mathrm{FM}_1} &= (1.000, 1.453, 1.207, 0.752, 1.818, 1.874)' \quad \text{for } g = 1, 2, \\
\widehat{\phi}_{1,1,1}^{\mathrm{FM}_1} &= 0.104, \\
\widehat{\phi}_{1,1,2}^{\mathrm{FM}_1} &= 0.345,
\end{aligned} \tag{16}$$

and those estimated under RG scaling are:

$$\begin{aligned}
\widehat{\boldsymbol{\Lambda}}_g^{\mathrm{RG}} &= (0.323, 0.470, 0.390, 0.243, 0.588, 0.606)' \quad \text{for } g = 1, 2, \\
\widehat{\phi}_{1,1,1}^{\mathrm{RG}} &= 1.000, \\
\widehat{\phi}_{1,1,2}^{\mathrm{RG}} &= 3.302.
\end{aligned} \tag{17}$$

Now, we choose $\sqrt{0.104}$ as constant, which is the square root of the estimated latent variance $\widehat{\phi}_{1,1,1}^{\mathrm{FM}_1}$, and apply the following conversion:

$$\begin{aligned}
\widehat{\boldsymbol{\Lambda}}_g^{\mathrm{RG}} &= \sqrt{0.104} \cdot \widehat{\boldsymbol{\Lambda}}_g^{\mathrm{FM}_1} = \sqrt{0.104} \cdot (1.000, 1.453, 1.207, 0.752, 1.818, 1.874) \quad \text{for } g = 1, 2, \\
&= (0.323, 0.470, 0.390, 0.243, 0.588, 0.606)', \\
\widehat{\phi}_{1,1,1}^{\mathrm{RG}} &= \frac{1}{\sqrt{0.104}^2} \cdot \widehat{\phi}_{1,1,1}^{\mathrm{FM}_1} = \frac{1}{\sqrt{0.104}^2} \cdot 0.104 = 1.000, \\
\widehat{\phi}_{1,1,2}^{\mathrm{RG}} &= \frac{1}{\sqrt{0.104}^2} \cdot \widehat{\phi}_{1,1,2}^{\mathrm{FM}_1} = \frac{1}{\sqrt{0.104}^2} \cdot 0.345 = 3.302.
\end{aligned} \tag{18}$$

The choice of the constant exemplifies the mechanism of the conversion: It yields an "estimated" latent variance in the reference group of 1, which corresponds to the restriction prescribed by the RG scaling method.

To convert from FM_1 to EC, we must bear in mind that the EC scaling restriction requires $\mathrm{A}(\widehat{\lambda}_{\bullet,1}^{\mathrm{EC}}) = 1$. The loadings estimated under EC scaling are:

$$\begin{aligned}
\widehat{\boldsymbol{\Lambda}}_g^{\mathrm{EC}} &= (0.740, 1.076, 0.893, 0.557, 1.346, 1.388)' \quad \text{for } g = 1, 2, \\
\widehat{\phi}_{1,1,1}^{\mathrm{EC}} &= 0.191, \\
\widehat{\phi}_{1,1,2}^{\mathrm{EC}} &= 0.629.
\end{aligned} \tag{19}$$

Denoting the average of the estimated loadings under the FM_1 scaling as $\mathrm{A}(\widehat{\lambda}^{\mathrm{FM}_1}_{\bullet,1})$, then choosing the constant $1/\mathrm{A}(\widehat{\lambda}^{\mathrm{FM}_1}_{\bullet,1}) = 1.351$ gives the conversion:[7]

$$\begin{aligned}
\widehat{\boldsymbol{\Lambda}}^{\mathrm{EC}}_g &= \frac{1}{1.351} \cdot (1.000, 1.453, 1.207, 0.752, 1.818, 1.874)' \quad \text{for } g = 1, 2, \\
&= (0.740, 1.076, 0.893, 0.557, 1.346, 1.388)', \\
\widehat{\phi}^{\mathrm{EC}}_{1,1,1} &= 1.351^2 \cdot 0.104 = 0.191, \\
\widehat{\phi}^{\mathrm{EC}}_{1,1,2} &= 1.351^2 \cdot 0.345 = 0.629.
\end{aligned} \tag{20}$$

Klopp and Klößner (2021, Proposition 2) provide a theoretical account that demonstrates how to convert the estimated parameters obtained under any scaling method to the estimates obtained under any other scaling method, without the need to re-estimate the model, given that the estimated model-implied covariance matrices under the two different scaling methods are identical. To start, we note that the scaling method that was actually used has no effect on the conversion, only the "target" scaling method is of relevance (Klopp & Klößner, 2021, p. 190). For our further considerations, let $S_1, S_2 \in \mathcal{S}$ be two different scaling methods. The model was actually estimated with S_1 scaling and S_2 is the target scaling method. Additionally, let $d \in \mathbb{R}\backslash\{0\}$ be a constant. For our one latent variable example, the conversion equations according to Klopp and Klößner (2021) are

$$\widehat{\lambda}^{S_2}_{j,i,g} = d \cdot \widehat{\lambda}^{S_1}_{j,i,g} \tag{21}$$

$$\widehat{\phi}^{S_2}_{i,i,g} = \frac{1}{d^2} \cdot \widehat{\phi}^{S_1}_{i,i,g} \tag{22}$$

Obviously, for our example with one latent variable, a conversion according to Equation (21) and Equation (22) with a constant d_1 does not change the estimated model-implied covariance matrix, because

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}^{S_2}_g &= \widehat{\boldsymbol{\Lambda}}^{S_2}_g \widehat{\boldsymbol{\Phi}}^{S_2}_g \widehat{\boldsymbol{\Lambda}}^{S_2\prime}_g + \widehat{\boldsymbol{\Theta}}_g = \\
&= d\widehat{\boldsymbol{\Lambda}}^{S_1}_g \left(\frac{1}{d^2} \widehat{\boldsymbol{\Phi}}^{S_1}_g \right) d\widehat{\boldsymbol{\Lambda}}^{S_1\prime}_g + \widehat{\boldsymbol{\Theta}}_g \\
&= \widehat{\boldsymbol{\Lambda}}^{S_1}_g \widehat{\boldsymbol{\Phi}}^{S_1}_g \widehat{\boldsymbol{\Lambda}}^{S_1\prime}_g + \widehat{\boldsymbol{\Theta}}_g = \widehat{\boldsymbol{\Sigma}}^{S_1}_g .
\end{aligned} \tag{23}$$

The type of relation in Equation (23) is widely known in the SEM literature (e.g., Jöreskog, 1978; Mulaik, 2010, p. 443; Yoon and Millsap, 2007); Klopp and Klößner (2021,

7) The exact number is 1.350715. Using the rounded number 1.351 produces some rounding issues in the following calculation.

p. 185) call this relation *change of scale*. Note that this relation deserves special attention in the multiple group context. Because of the invariance conditions over the two groups, only one constant must be used simultaneously for both groups.

A general version of a change of scale for the model in Equation (9), which also easily accommodates more than one latent variable, would use non-singular diagonal matrices $\boldsymbol{D}$ containing the d_i factors for each latent variable. This general form is

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_g^{S_2} &= \widehat{\boldsymbol{\Lambda}}_g^{S_2}\widehat{\boldsymbol{\Phi}}_g^{S_2}\widehat{\boldsymbol{\Lambda}}_g^{S_2\prime} + \widehat{\boldsymbol{\Theta}}_g = \\ &= \widehat{\boldsymbol{\Lambda}}_g^{S_1}\boldsymbol{D}\left(\boldsymbol{D}^{-1}\widehat{\boldsymbol{\Phi}}_g^{S_1}\boldsymbol{D}^{-1}\right)\widehat{\boldsymbol{\Lambda}}_g^{S_1\prime}\boldsymbol{D}' + \widehat{\boldsymbol{\Theta}}_g \\ &= \widehat{\boldsymbol{\Lambda}}_g^{S_1}\boldsymbol{D}\boldsymbol{D}^{-1}\widehat{\boldsymbol{\Phi}}_g^{S_1}\boldsymbol{D}^{-1}\boldsymbol{D}\widehat{\boldsymbol{\Lambda}}_g^{S_1\prime} + \widehat{\boldsymbol{\Theta}}_g \\ &= \widehat{\boldsymbol{\Lambda}}_g^{S_1}\widehat{\boldsymbol{\Phi}}_g^{S_1}\widehat{\boldsymbol{\Lambda}}_g^{S_1\prime} + \widehat{\boldsymbol{\Theta}}_g = \widehat{\boldsymbol{\Sigma}}_g^{S_1}\,.\end{aligned} \tag{24}$$

In the following section, we turn over to the formal side of this exemplary consideration and provide a general proof that each scaling method results in the same value of the discrepancy function. As we will see, the idea of a change of scale and the idea of the propagation of the scaling restriction over the groups via the invariance conditions will be essential in proving the equivalence of the scaling methods.

Full Metric Measurement Invariance: Theory

The considerations laid out above lead to the following proposition.

Proposition 1: If the full metric MI model is estimated by minimizing a discrepancy function, then the resulting optimal values of the discrepancy function as well as the estimated model-implied covariance matrices do not depend on the particular method used for scaling the MI model.

The outline of the proof is as follows: First of all, we will explain why the minimum of the discrepancy function, taken over all parameters fulfilling full metric MI, coincides, for every scaling method, with the discrepancy function's minimum taken over all parameters simultaneously fulfilling full metric MI as well as the restrictions stemming from the corresponding scaling method. After having established this fact, it will be obvious that the resulting optimal discrepancy value does not depend on the particular scaling method, as the optimal values for different scaling methods all take the same value (namely, the discrepancy function's minimum taken over all parameters fulfilling full metric MI). As a by-product of proving the invariance of the optimal discrepancy value, the proof will also show that the estimated model-implied covariance matrices do not depend on the method used for scaling the MI model.

To follow the outline described above, we denote by $\widehat{\theta}$ an arbitrary parameter which fulfills full metric MI and minimizes the discrepancy function when no scaling restrictions are imposed. By definition, $F(\widehat{\theta}) \leq F(\theta)$ for all parameters θ which fulfill full metric MI, with F denoting the discrepancy function. For all scaling methods, this in particular implies $F(\widehat{\theta}) \leq F(\widehat{\theta}^S)$, with $\widehat{\theta}^S$ denoting the estimate for the parameters for a given scaling method S. We will now show that the reverse relation holds, too, i.e. $F(\widehat{\theta}) \geq F(\widehat{\theta}^S)$. When this relation has been established, it is clear that we have equality, $F(\widehat{\theta}) = F(\widehat{\theta}^S)$, implying that all scaling methods lead to the same value for discrepancy function, $F(\widehat{\theta})$, for the full metric MI model. To this end, we show that, for each scaling method S, there exists a change of scale which transforms the parameter $\widehat{\theta}$ to a new parameter $\widetilde{\theta}^S$, such that the transformed parameter $\widetilde{\theta}^S$ fulfills the constraints associated with the scaling method S as well as full metric MI.

1. For the FM method, we apply to $\widehat{\theta}$ the change of scale which divides all loadings of indicators j belonging to latent variable i in group g by $\widehat{\lambda}_{R_i, i, 1}$, where R_i denotes the RI used for scaling factor i, while multiplying latent covariances $\widehat{\phi}_{i_1, i_2, g}$ between latent variables i_1 and i_2 in group g by $\widehat{\lambda}_{R_{i_1}, i_1, 1} \cdot \widehat{\lambda}_{R_{i_2}, i_2, 1}$.[8] This transformation results in a new parameter for which the RI's loading always equals 1 (due to $\widetilde{\lambda}^S_{R_i, i, g} = \frac{\widehat{\lambda}_{R_i, i, g}}{\widehat{\lambda}_{R_i, i, 1}} = \frac{\widehat{\lambda}_{R_i, i, 1}}{\widehat{\lambda}_{R_i, i, 1}} = 1$) and all indicators are invariant (as the already invariant loadings of $\widehat{\theta}$ are divided by a quantity which is also invariant across groups). Thus, the transformed parameter fulfills the constraints associated with FM scaling as well as full metric MI.
2. For the EC method, we apply to $\widehat{\theta}$ the change of scale which divides all loadings of indicators j belonging to latent variable i by $\mathrm{A}(\widehat{\lambda}_{\bullet, i, 1})$, the average of all the loadings of latent variable i in group 1, while multiplying latent covariances $\widehat{\phi}_{i_1, i_2, g}$ between latent variables i_1 and i_2 in group g by $\mathrm{A}(\widehat{\lambda}_{\bullet, i_1, 1}) \cdot \mathrm{A}(\widehat{\lambda}_{\bullet, i_2, 1})$. This transformation results in a new parameter for which the average loading of latent variable i's indicators always equals 1 (due to $\mathrm{A}(\widetilde{\lambda}^S_{\bullet, i, g}) = \mathrm{A}(\frac{\widehat{\lambda}_{\bullet, i, g}}{\mathrm{A}(\widehat{\lambda}_{\bullet, i, 1})}) = \mathrm{A}(\frac{\widehat{\lambda}_{\bullet, i, 1}}{\mathrm{A}(\widehat{\lambda}_{\bullet, i, 1})}) = 1$) and all indicators are invariant (as the already invariant loadings of $\widehat{\theta}$ are divided by a quantity which is also invariant across groups). Thus, the transformed parameter fulfills the constraints associated with EC scaling as well as full metric MI.
3. For the RG method, we apply to $\widehat{\theta}$ the change of scale which multiplies all loadings of indicators j belonging to latent variable i by $\sqrt{\widehat{\phi}_{i, i, g_0}}$, latent variable i's standard

8) Latent variances, being the special case of covariances with $i_1 = i_2$, are thus multiplied by $\widehat{\lambda}^2_{R_{i_1}, i_1, 1}$.

deviation in the reference group g_0, while dividing latent covariances $\widehat{\phi}_{i_1, i_2, g}$ between latent variables i_1 and i_2 in group g by $\sqrt{\widehat{\phi}_{i_1, i_1, g_0}} \cdot \sqrt{\widehat{\phi}_{i_2, i_2, g_0}}$. This transformation results in a new parameter for which the latent variable's variance in group g_0 equals 1 (due to $\widetilde{\phi}^S_{i, i, g_0} = \frac{\widehat{\phi}_{i, i, g_0}}{\sqrt{\widehat{\phi}_{i, i, g_0}} \cdot \sqrt{\widehat{\phi}_{i, i, g_0}}} = 1$) and all indicators are invariant (as the already invariant loadings of $\widehat{\theta}$ are multiplied by a quantity which is also invariant across groups). Thus, the transformed parameter fulfills the constraints associated with RG scaling as well as full metric MI.

In all three cases, the change of scale can be described by $\widetilde{\boldsymbol{\Lambda}}^S_g = \widehat{\boldsymbol{\Lambda}}_g \boldsymbol{D}$ and $\widetilde{\boldsymbol{\Phi}}^S_g = \boldsymbol{D}^{-1} \widehat{\boldsymbol{\Phi}}_g \boldsymbol{D}^{-1}$ for all groups g, where $\boldsymbol{D}$ is a diagonal matrix (with entries $d_{i,i} = \frac{1}{\widehat{\lambda}_{R_g, i, 1}}$ for FM, $d_{i,i} = \frac{1}{\mathrm{A}(\widehat{\lambda}_{\bullet, i, 1})}$ for EC, and $d_{i,i} = \sqrt{\widehat{\phi}_{i, i, g_0}}$ for RG scaling). From

$$\widetilde{\boldsymbol{\Lambda}}^S_g \widetilde{\boldsymbol{\Phi}}^S_g \widetilde{\boldsymbol{\Lambda}}^{S'}_g = \widehat{\boldsymbol{\Lambda}}_g \boldsymbol{D}\, \boldsymbol{D}^{-1} \widehat{\boldsymbol{\Phi}}_g \boldsymbol{D}^{-1} \left(\widehat{\boldsymbol{\Lambda}}_g \boldsymbol{D}\right)' = \widehat{\boldsymbol{\Lambda}}_g \widehat{\boldsymbol{\Phi}}_g \boldsymbol{D}^{-1} \boldsymbol{D} \widehat{\boldsymbol{\Lambda}}'_g = \widehat{\boldsymbol{\Lambda}}_g \widehat{\boldsymbol{\Phi}}_g \widehat{\boldsymbol{\Lambda}}'_g \tag{25}$$

it follows that, in all groups, the estimated model-implied covariance matrices are identical for $\widetilde{\theta}^S$ and $\widehat{\theta}$. As the discrepancy function depends on the parameters only through the estimated model-implied covariance matrices, we can conclude that $F(\widehat{\theta}) = F(\widetilde{\theta}^S) \geq F(\widehat{\theta}^S)$, as $\widehat{\theta}^S$ minimizes the discrepancy function among all parameters which fulfill both full metric MI and the restrictions coming with the scaling method S, a set to which $\widetilde{\theta}^S$ belongs.[9] Additionally, Equation (25) shows that the estimated model-implied covariance matrices do not depend on the method used for scaling the full metric MI model.

Corollary 1: If the full metric MI model is estimated by minimizing a discrepancy function, then its χ^2, p value, RMSEA, and other fit measures do not depend on the particular method used for scaling the model.

This immediately follows from Proposition 1, as all these quantities are calculated using the full metric MI model's likelihood value, which does not depend on the particular method used for scaling the model.

Corollary 2: The results of the χ^2 difference test which compares the full metric MI model to the configural MI model do not depend on the scaling methods used for estimating the configural and full metric MI model. This in particular holds for the χ^2 difference statistics, the p value, ΔRMSEA, and differences of other fit measures.

9) If the estimate under the scaling method is uniquely defined, then $\widehat{\theta}^S$ and $\widetilde{\theta}^S$ will coincide, $\widehat{\theta}^S = \widetilde{\theta}^S$.

This immediately follows from the preceding corollary and the well-known fact that the corresponding quantities for the configural model do not depend on the particular method used for scaling the configural model, either.

Concerning the χ^2-difference test comparing the configural to the full metric MI model, we would also like to shed some light on how to determine its degrees of freedom. Our explanations above show that scaling the full metric MI model consists of imposing exactly m restrictions, for instance by fixing all latent variables' variances to 1 in one group, as it is the case when scaling is done using the RG method. On the other hand, it is well-known that scaling the configural model consists of imposing exactly Gm restrictions, namely one per group and latent variable. Thus, there is a difference of $Gm - m = (G-1)m$ between the numbers of restrictions needed to scale the configural model and the full metric MI model, respectively. This difference needs to be taken into account when calculating the degrees of freedom of the χ^2-difference test. Intuitively, one would think that the degrees of freedom were determined by the number of restrictions due to the full metric MI model, i.e. by the number of conditions required to ensure all loadings' invariance. This number is given by $(G-1)p$, as, for every indicator, the loadings in the first group must match the loadings in the remaining $G-1$ groups. However, this number has to be corrected to account for the difference induced by the scaling restrictions, i.e. by $(G-1)m$. These considerations yield the following corollary:

Corollary 3: The degrees of freedom of the χ^2-difference test between the configural and full metric MI model are given by:

$$df_{\text{full}} - df_{\text{configural}} = \Delta df = (G-1)p - (G-1)m = (G-1)(p-m)\,.$$

Corollary 3 can easily be adapted to partial metric MI models. In this case, the overall number of invariant indicators will be denoted by p^*. To calculate the degrees of freedom for the χ^2 difference test comparing the partial metric MI model to the configural one, we can apply exactly the same approach as above: the number of restrictions due to the partial metric MI model, intuitively given by $(G-1)p^*$, has to be corrected to account for the difference induced by the scaling restrictions, i.e. by $(G-1)m$. This results in the following corollary:

Corollary 4: The degrees of freedom of the χ^2-difference test between the configural and a partial metric MI model are given by:

$$df_{\text{partial}} - df_{\text{configural}} = \Delta df = (G-1)p^* - (G-1)m = (G-1)(p^*-m)\,.$$

The preceding corollary implies a special case which deserves particular attention: the case of $p^* = m$, which leads to the χ^2-difference test having zero degrees of freedom. Regarding this case, we first we need to investigate when it appears: the condition $p^* = m$ means that the number of indicators required to have invariant loadings, p^*, equals the

number of latent variables, m. This happens if exactly one indicator per latent variable is presumed to be invariant. Regarding the consequences, the fact that the χ^2-difference test has zero degrees of freedom implies that these particular partial metric MI models have as many degrees of freedom as the configural model. Even more, as partial metric MI models are covariance-nested within the configural model (cf., Bentler & Bonett, 1980, p. 592f), all these partial metric MI models are indeed equivalent to the configural model. Together, this leads to the following corollary:

Corollary 5: A partial metric MI model in which only one indicator per latent variable is presumed to be invariant is equivalent to the configural model.

Partial Metric Measurement Invariance: An Example

Up to this point, we have mainly considered full metric MI models. In this section, building on the corollaries developed in the preceding section, we want to present a further example, in which we look at partial metric MI models. The example consists of three scenarios, A, B, and C. Each of thee scenarios is further divided into two settings. The dgp stays the same as in the previous example, please see the model given in Equation (10). As before, we use the ML discrepancy function and take the data from the first sample of the previous example's simulation.

In the first scenario A, we want to look at the special case with an invariance condition on the estimated loading of only one indicator. In this case, there is $\Delta df = 0$ and the model under consideration is equivalent to the configural model, which has $df_{\text{configural}} = 18$ degrees of freedom (please see the supporting information for how the degrees of freedom are calculated). Estimating the configural model[10] results in a value of $19404 \cdot 10^{-6}$ for the discrepancy function.

In the first setting of scenario A, we examine the case in which the estimated loading of the first indicator, which is actually invariant according to the dgp, is assumed to be invariant by the researcher. Thus, the unscaled model is identical to Equation (5) with $g = 1, 2$ and with the metric invariance condition $\widehat{\lambda}_{1,1,1} = \widehat{\lambda}_{1,1,2}$. To scale this model, we start with the FM scaling method, see Table 3 for details.[11] Once again, we consider every indicator as a potential RI, i.e., we look at $\text{FM}_1, \ldots, \text{FM}_6$. For the FM_1 scaling, the invariance condition $\widehat{\lambda}_{1,1,1} = \widehat{\lambda}_{1,1,2}$ yields that estimated loadings equal 1 in both groups. For the FM_2 to FM_6 scaling, the invariance condition regarding the first loading propagates

10) We used the FM_1 scaling method in each group. Because the configural model is basically a CFA model estimated simultaneously in several groups, the scaling has to be done in each group, and in each group the value of the discrepancy function does not depend on the scaling method.

11) Again, we only consider the case when the scaling restrictions are set in the reference group.

the scaling from the reference to the focal group. Concerning the EC method, a first variant is to constrain the sum of only invariant indicators' estimated loadings, which is for instance the way the semTools package (Jorgensen et al., 2021) applies this scaling method in the context of partial metric MI models.[12] In this case, this leads to restricting the estimated loading of the first indicator to 1 (EC_1). Thus, EC_1 coincides with the FM_1 scaling in this case. A second possibility is to constrain the sum of the estimated loadings of indicators which are not presumed to be invariant, such that their sum equals 5 (EC_2). As a third possibility, we might also incorporate all estimated loadings, regardless of them being invariant or not, resulting in the condition that the sum of all estimated loadings equals 6 (EC_3). Finally, there is the RG scaling method (RG_1).

The second setting of scenario A entails an invariance condition on the estimated loading of the fifth indicator, i.e., $\widehat{\lambda}_{5,1,1} = \widehat{\lambda}_{5,1,2}$, which is non-invariant in the dgp. The scaling options are basically the same as in the first setting, they differ only with respect to which indicators the researcher deems invariant or non-invariant, see Table 3 for details.

Estimating the models produces the results shown in Table 4. The table shows that in both settings, the value of the discrepancy function multiplied by the factor 10^6 is 19404. Thus, estimation leads to the same value of the discrepancy function, although there are different invariance conditions, and these partial models are indeed equivalent to the configural model. The scenario shows exemplarily that a partial metric MI condition on only one estimated loading is not feasible. Scrutinizing the various scaling options in Table 3 provides insight into this issue. For example, the FM_1 scaling in the first setting as well as the FM_5 scaling in the second setting correspond to the respective FM scaling of the configural model, where the RI's loading is set to 1 in *both* groups. The reason is the propagation of the scaling via the invariance condition. Interestingly, this way of looking at the configural model is not new: Reise et al. (1993) developed a scaling method for the configural model, in which they set the variance of the latent variable to 1 in one group and required one loading to be invariant across groups. Their scaling method for the configural model is therefore identical to the RG scaling method for a partial metric MI model with only one invariant indicator. Steenkamp and Baumgartner (1998, p. 81) also pointed out that testing the invariance of only one loading is not meaningful. Following from our results, it is impossible to test the invariance of only one indicator because such a model will always correspond to the configural model.

In order to get around the problem arising in Scenario A, one might be tempted to ignore the rule of applying scaling restrictions in only one group and apply them in both groups. Doing so, there would be one degree of freedom for the χ^2-difference test, and the partial metric MI model would no longer be equivalent to the configural model. In

12) This refers to the function measEq.syntax in the semTools package.

Table 3

Invariance Conditions and Scaling Restrictions for All Settings in Scenario A in the Partial Metric MI Example

Setting 1	
Invariance condition	$\widehat{\lambda}_{1,1,1} = \widehat{\lambda}_{1,1,2}$
Scaling	$\mathrm{FM}_1: \quad \widehat{\lambda}_{1,1,1} = 1$
	$\mathrm{FM}_2: \quad \widehat{\lambda}_{2,1,1} = 1$
	$\mathrm{FM}_3: \quad \widehat{\lambda}_{3,1,1} = 1$
	$\mathrm{FM}_4: \quad \widehat{\lambda}_{4,1,1} = 1$
	$\mathrm{FM}_5: \quad \widehat{\lambda}_{5,1,1} = 1$
	$\mathrm{FM}_6: \quad \widehat{\lambda}_{6,1,1} = 1$
	$\mathrm{EC}_1: \quad \widehat{\lambda}_{1,1,1} = 1$
	$\mathrm{EC}_2: \quad \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 5$
	$\mathrm{EC}_3: \quad \widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 6$
	$\mathrm{RG}_1: \quad \widehat{\phi}_{1,1,1} = 1$
Setting 2	
Invariance condition	$\widehat{\lambda}_{5,1,1} = \widehat{\lambda}_{5,1,2}$
Scaling	$\mathrm{FM}_1: \quad \widehat{\lambda}_{1,1,1} = 1$
	$\mathrm{FM}_2: \quad \widehat{\lambda}_{2,1,1} = 1$
	$\mathrm{FM}_3: \quad \widehat{\lambda}_{3,1,1} = 1$
	$\mathrm{FM}_4: \quad \widehat{\lambda}_{4,1,1} = 1$
	$\mathrm{FM}_5: \quad \widehat{\lambda}_{5,1,1} = 1$
	$\mathrm{FM}_6: \quad \widehat{\lambda}_{6,1,1} = 1$
	$\mathrm{EC}_1: \quad \widehat{\lambda}_{5,1,1} = 1$
	$\mathrm{EC}_2: \quad \widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{6,1,1} = 5$
	$\mathrm{EC}_3: \quad \widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 6$
	$\mathrm{RG}_1: \quad \widehat{\phi}_{1,1,1} = 1$

particular, such an approach would lead to a model with $df = 19$ degrees of freedom, but at the cost of violating the rule to apply the scaling restriction in only one group.[13]

13) Such a situation may happen when there is a configural model, which has a scaling restriction in each group, and a researcher sets an invariance condition on an estimated loading falsely believing that this *were* a partial metric MI model with one invariance condition.

Table 4

Values of the ML Discrepancy Function Multiplied by the Factor 10^6 in Scenario A in the Partial Metric MI Example

Setting	Discrepancy function values							
1	FM_2	FM_3	FM_4	FM_5	FM_6	EC_1	EC_2	RG_1
	19404	19404	19404	19404	19404	19404	19404	19404
2	FM_1	FM_2	FM_3	FM_4	FM_6	EC_1	EC_2	RG_1
	19404	19404	19404	19404	19404	19404	19404	19404

In the following Scenario B, we will study this situation. The unscaled models in the two settings are identical to those in Scenario A. The various scaling options are shown in Table 5. Please note that in the first setting, FM_1 scaling coincides with the configural model, as restricting the first loading in each group is the same as imposing an invariance condition. An analogous assertion holds for FM_5 scaling in the second setting. The resulting values for the ML discrepancy function are given in Table 6. Somewhat surprisingly, almost all scaling methods yield different values for the discrepancy function (for the exception EC scaling we will provide hints below). Thus, the model fit as well as test statistics for the χ^2-difference test are idiosyncratic to each scaling method, a phenomenon called *constraint interaction* (Klößner & Klopp, 2019; Steiger, 2002). The reason is that enforcing scaling restrictions in both groups leads to the hypothesis about the indicators' invariance becoming dependent on the scaling method employed. The findings of Johnson et al. (2009) mentioned in the introduction are also driven by this effect, which we will now explore in more depth.

To do so, let us have a look at the situation in Setting 1 and the use of FM_2 to scale the model. In this model, the first indicator's estimated loading is invariant due to the condition $\widehat{\lambda}_{1,1,1} = \widehat{\lambda}_{1,1,2}$, while the second indicator's loading indirectly becomes invariant due to the scaling restrictions $\widehat{\lambda}_{2,1,1} = 1$ and $\widehat{\lambda}_{2,1,2} = 1$, which obviously imply $\widehat{\lambda}_{2,1,1} = \widehat{\lambda}_{2,1,2}$. Therefore, this model does not presume that only the first indicator's loading is invariant, but it implicitly stipulates that the first and second indicators' loadings are both invariant. Indeed, as the conditions $\widehat{\lambda}_{2,1,1} = 1$ and $\widehat{\lambda}_{2,1,2} = 1$ are equivalent to the conditions $\widehat{\lambda}_{2,1,1} = \widehat{\lambda}_{2,1,2}$ and $\widehat{\lambda}_{2,1,1} = 1$, the FM_2-scaled model in the first setting is actually equivalent to a partial metric MI model where the first two indicators' loadings are presumed to be invariant and FM_2 scaling is used in the first group only. Thus, the FM_2-scaled model in the first setting is a model with $p^* = 2$, not $p^* = 1$, invariance conditions, investigating whether the first two indicators' loadings are invariant across groups.

Scenario B demonstrates another aspect, too. There may be conditions under which identical values of the discrepancy function emerge, even though the model is not scaled

Table 5

Invariance Conditions and Scaling Restrictions for All Settings in Scenario B in the Partial Metric MI Example

Setting 1	
Invariance condition	$\widehat{\lambda}_{1,1,1} = \widehat{\lambda}_{1,1,2}$
Scaling	$\mathrm{FM}_2: \ \widehat{\lambda}_{2,1,1} = 1 \wedge \widehat{\lambda}_{2,1,2} = 1$
	$\mathrm{FM}_3: \ \widehat{\lambda}_{3,1,1} = 1 \wedge \widehat{\lambda}_{3,1,2} = 1$
	$\mathrm{FM}_4: \ \widehat{\lambda}_{4,1,1} = 1 \wedge \widehat{\lambda}_{4,1,2} = 1$
	$\mathrm{FM}_5: \ \widehat{\lambda}_{5,1,1} = 1 \wedge \widehat{\lambda}_{5,1,2} = 1$
	$\mathrm{FM}_6: \ \widehat{\lambda}_{6,1,1} = 1 \wedge \widehat{\lambda}_{6,1,2} = 1$
	$\mathrm{EC}_1: \ \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 5 \ \wedge$
	$\widehat{\lambda}_{2,1,2} + \widehat{\lambda}_{3,1,2} + \widehat{\lambda}_{4,1,2} + \widehat{\lambda}_{5,1,2} + \widehat{\lambda}_{6,1,2} = 5$
	$\mathrm{EC}_2: \ \widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 6 \ \wedge$
	$\widehat{\lambda}_{1,1,2} + \widehat{\lambda}_{2,1,2} + \widehat{\lambda}_{3,1,2} + \widehat{\lambda}_{4,1,2} + \widehat{\lambda}_{5,1,2} + \widehat{\lambda}_{6,1,2} = 6$
	$\mathrm{RG}: \ \widehat{\phi}_{1,1,1} = 1 \wedge \widehat{\phi}_{1,1,2} = 1$
Setting 2	
Invariance condition	$\widehat{\lambda}_{5,1,1} = \widehat{\lambda}_{5,1,2}$
Scaling	$\mathrm{FM}_1: \ \widehat{\lambda}_{1,1,1} = 1 \wedge \widehat{\lambda}_{1,1,2} = 1$
	$\mathrm{FM}_2: \ \widehat{\lambda}_{2,1,1} = 1 \wedge \widehat{\lambda}_{2,1,2} = 1$
	$\mathrm{FM}_3: \ \widehat{\lambda}_{3,1,1} = 1 \wedge \widehat{\lambda}_{3,1,2} = 1$
	$\mathrm{FM}_4: \ \widehat{\lambda}_{4,1,1} = 1 \wedge \widehat{\lambda}_{4,1,2} = 1$
	$\mathrm{FM}_6: \ \widehat{\lambda}_{6,1,1} = 1 \wedge \widehat{\lambda}_{6,1,2} = 1$
	$\mathrm{EC}_1: \ \widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{6,1,1} = 5 \ \wedge$
	$\widehat{\lambda}_{1,1,2} + \widehat{\lambda}_{2,1,2} + \widehat{\lambda}_{3,1,2} + \widehat{\lambda}_{4,1,2} + \widehat{\lambda}_{6,1,2} = 5$
	$\mathrm{EC}_2: \ \widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 6 \ \wedge$
	$\widehat{\lambda}_{1,1,2} + \widehat{\lambda}_{2,1,2} + \widehat{\lambda}_{3,1,2} + \widehat{\lambda}_{4,1,2} + \widehat{\lambda}_{5,1,2} + \widehat{\lambda}_{6,1,2} = 6$
	$\mathrm{RG}: \ \widehat{\phi}_{1,1,1} = 1 \wedge \widehat{\phi}_{1,1,2} = 1$

according to the rule of setting scaling restrictions in one group only, as it was the case for the EC scaling variants in this scenario[14] .

14) The reason for this are specific to the EC scaling, a detailed explanation can be found in the supporting information.

Table 6

Values of the ML Discrepancy Function Multiplied by the Factor 10^6 in Scenario B in the Partial Metric MI Example

Setting	Discrepancy function values							
1	FM_2	FM_3	FM_4	FM_5	FM_6	EC_1	EC_2	RG_1
	19863	19469	20673	24692	35574	23825	23825	28653
2	FM_1	FM_2	FM_3	FM_4	FM_6	EC_1	EC_2	RG_1
	20647	23324	28741	19706	24553	21314	21314	71042

We now consider the last scenario, Scenario C, in which we want to showcase the scaling of a partial metric MI model with $p^* = 4$ indicators presumed to be invariant across groups and with obeying the rule of applying scaling restrictions in only one group. Consequently, this partial metric MI model has $df_{\text{partial}} = 21$ degrees of freedom. *The first setting* represents a situation in which the first four estimated loadings are investigated with respect to invariance across groups, i.e., the invariance conditions stipulated by the researcher correspond exactly to the invariant indicators in the dgp. Again, the unscaled model is as given in Equation (5) with $g = 1, 2$, and with the metric invariance conditions $\widehat{\lambda}_{j,1,1} = \widehat{\lambda}_{j,1,2}$ for $j = 1, 2, 3, 4$.

For scaling, we use all possible versions of the FM, EC, and RG scaling methods. For the FM scaling method, these are six different versions, i.e., $FM_1, \ldots, FM_6$. For indicators with invariance conditions, we see that if we set the FM scaling restriction in the reference group, the invariance conditions automatically propagate the corresponding condition to the focal group. For instance, for the FM_1 scaling, the restriction $\widehat{\lambda}_{1,1,1} = 1$ is propagated to the focal group via the invariance conditions $\widehat{\lambda}_{1,1,1} = \widehat{\lambda}_{1,1,2}$. This reasoning applies for FM_2, FM_3, and FM_4 scaling, too. For FM_5 and FM_6 scaling, where the reasoning given above no longer applies, the partial metric invariance conditions on the first four loadings nevertheless propagate the scaling from the reference to the focal group.

For the EC scaling method, there are again several ways to set the restriction. The first way (EC_1, see Table 7), consists in restricting the invariant loadings such that their sum equals 4, while the second way leads to restricting the two loadings without invariance conditions (EC_2). Finally, it is also possible to apply the scaling restrictions such that the sum of all loadings equals 6 (EC_3). As above, these restrictions are applied in the first group only, following our rule that the scaling restrictions *must not* be applied in both groups.

Table 7

Invariance Conditions and Scaling Restrictions for all Settings in Scenario C in the Partial Metric MI Example.

Setting 1	
Invariance conditions	$\widehat{\lambda}_{1,1,1} = \widehat{\lambda}_{1,1,2}$, $\widehat{\lambda}_{2,1,1} = \widehat{\lambda}_{2,1,2}$, $\widehat{\lambda}_{3,1,1} = \widehat{\lambda}_{3,1,2}$, $\widehat{\lambda}_{4,1,1} = \widehat{\lambda}_{4,1,2}$
Scaling	$\mathrm{FM}_1: \quad \widehat{\lambda}_{1,1,1} = 1$
	$\mathrm{FM}_2: \quad \widehat{\lambda}_{2,1,1} = 1$
	$\mathrm{FM}_3: \quad \widehat{\lambda}_{3,1,1} = 1$
	$\mathrm{FM}_4: \quad \widehat{\lambda}_{4,1,1} = 1$
	$\mathrm{FM}_5: \quad \widehat{\lambda}_{5,1,1} = 1$
	$\mathrm{FM}_6: \quad \widehat{\lambda}_{6,1,1} = 1$
	$\mathrm{EC}_1: \quad \widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} = 4$
	$\mathrm{EC}_2: \quad \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 2$
	$\mathrm{EC}_3: \quad \widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 6$
	$\mathrm{RG}_1: \quad \widehat{\phi}_{1,1,1} = 1$
Setting 2	
Invariance conditions	$\widehat{\lambda}_{1,1,1} = \widehat{\lambda}_{1,1,2}$, $\widehat{\lambda}_{2,1,1} = \widehat{\lambda}_{2,1,2}$, $\widehat{\lambda}_{5,1,1} = \widehat{\lambda}_{5,1,2}$, $\widehat{\lambda}_{6,1,1} = \widehat{\lambda}_{6,1,2}$
Scaling	$\mathrm{FM}_1: \quad \widehat{\lambda}_{1,1,1} = 1$
	$\mathrm{FM}_2: \quad \widehat{\lambda}_{2,1,1} = 1$
	$\mathrm{FM}_3: \quad \widehat{\lambda}_{3,1,1} = 1$
	$\mathrm{FM}_4: \quad \widehat{\lambda}_{4,1,1} = 1$
	$\mathrm{FM}_5: \quad \widehat{\lambda}_{5,1,1} = 1$
	$\mathrm{FM}_6: \quad \widehat{\lambda}_{6,1,1} = 1$
	$\mathrm{EC}_1: \quad \widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 4$
	$\mathrm{EC}_2: \quad \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} = 2$
	$\mathrm{EC}_3: \quad \widehat{\lambda}_{1,1,1} + \widehat{\lambda}_{2,1,1} + \widehat{\lambda}_{3,1,1} + \widehat{\lambda}_{4,1,1} + \widehat{\lambda}_{5,1,1} + \widehat{\lambda}_{6,1,1} = 6$
	$\mathrm{RG}_1: \quad \widehat{\phi}_{1,1,1} = 1$

Finally, for the RG scaling method, we set the variance of the latent variable in the reference group to 1. In contrast to the FM and EC scaling method, it is obvious for the RG scaling method that the scaling restriction is only applied in one group, as this feature is part of the description of this scaling method, as outlined in the introduction. The invariance conditions placed on the first four indicators certainly propagate the scaling to the focal group also for the EC and RG scaling.

In the second setting, there are four indicators again with invariance conditions on the estimated loadings. However, in this setting, the invariance conditions are placed on the

estimated loadings of the first and second indicator, which are invariant in the dgp, and on the estimated loadings of the fifth and sixth indicator, which are non-invariant in the dgp. Thus, the metric invariance conditions are $\widehat{\lambda}_{j,1,1} = \widehat{\lambda}_{j,1,2}$ for $j = 1, 2, 5, 6$. The scaling options are basically the same as in the first setting, they differ only with respect to which indicators the researcher deems invariant or non-invariant, see Table 7 for details. We would like to draw the reader's attention to the fact that the scaling methods follow the researcher's choices of invariant and non-invariant indicators, irrespective of the indicators' invariance or non-invariance in the dgp, which is unknown to the researcher.

The results are presented in Table 8, which shows the values of the ML discrepancy function multiplied by the factor 10^6. Obviously, the value of the discrepancy function, and therefore the test statistic and all other fit indices are identical, irrespective of the scaling method. Additionally, the example also demonstrates that the invariance or non-invariance of certain loadings in the dgp does not interact with the specific scaling method the researcher chooses for scaling the partial metric MI model.

Table 8

Values of the ML Discrepancy Function Multiplied by the Factor 10^6 in Scenario C in the Partial Metric MI Example

Setting	Discrepancy function values									
1	FM_1	FM_2	FM_3	FM_4	FM_5	FM_6	EC_1	EC_2	EC_3	RG_1
	22021	22021	22021	22021	22021	22021	22021	22021	22021	22021
2	FM_1	FM_2	FM_3	FM_4	FM_5	FM_6	EC_1	EC_2	EC_3	RG_1
	45195	45195	45195	45195	45195	45195	45195	45195	45195	45195

In the following section, we will provide the formal proof that for partial metric MI models, the scaling method does not affect the value of the discrepancy function, as long as scaling restrictions are applied in one group only.

Partial Metric Measurement Invariance: Theory

The following proposition formalizes the results of the previous section.

Proposition 2: If a partial metric MI model is estimated by minimizing a discrepancy function, then the resulting optimal values of the discrepancy function as well as the estimated model-implied covariance matrices do not depend on the particular method used for scaling the partial metric MI model, as long as the scaling method is applied in only one group, g_0.

The proof is essentially the same as the one for Proposition 1. The only detail that needs to be given precisely is how exactly to construct the changes of scale that transform a

given parameter $\widehat{\theta}$ to a new parameter $\widetilde{\theta}^S$ which fulfills the restrictions associated with the Scaling Method *S*.

1. For the FM method, where the scaling restrictions consist of latent variable i's (potentially non-invariant) reference indicator R_i's loading being 1 in Group g_0, we apply to $\widehat{\theta}$ the change of scale which divides all loadings of indicators j belonging to latent variable i in Group g by $\widehat{\lambda}_{R_i, i, g_0}$, while multiplying latent covariances $\widehat{\phi}_{i_1, i_2, g}$ between latent variables i_1 and i_2 in Group g by $\widehat{\lambda}_{R_{i_1}, i_1, g_0} \cdot \widehat{\lambda}_{R_{i_2}, i_2, g_0}$. This transformation results in a new parameter for which the reference indicator's loading in Group g_0 equals 1 (due to $\widetilde{\lambda}^S_{R_i, i, g_0} = \frac{\widehat{\lambda}_{R_i, i, g_0}}{\widehat{\lambda}_{R_i, i, g_0}} = 1$) and all indicators supposed to be invariant across groups stay invariant (as the already invariant loadings of $\widehat{\theta}$ are divided by a quantity which is also invariant across groups). Thus, the transformed parameter fulfills the constraints associated with FM scaling as well as those of the partial metric MI model.
2. For the EC method, we apply to $\widehat{\theta}$ the change of scale which divides all loadings of indicators j belonging to latent variable i by $A(\widehat{\lambda}_{\bullet, i, g_0})$, the average of the corresponding loadings of latent variable i in Group g_0, while multiplying latent covariances $\widehat{\phi}_{i_1, i_2, g}$ between latent variables i_1 and i_2 in Group g by $A(\widehat{\lambda}_{\bullet, i_1, g_0}) \cdot A(\widehat{\lambda}_{\bullet, i_2, g_0})$. This transformation results in a new parameter for which the average loading of the latent variable i's indicators in Group g_0 equals 1 (due to $A(\widetilde{\lambda}^S_{\bullet, i, g_0}) = A(\frac{\widehat{\lambda}_{\bullet, i, g_0}}{A(\widehat{\lambda}_{\bullet, i, g_0})}) = 1$) and all indicators supposed to be invariant across groups stay invariant (as the already invariant loadings of $\widehat{\theta}$ are divided by a quantity which is also invariant across groups). Thus, the transformed parameter fulfills the constraints associated with EC scaling as well as those of the partial metric MI model.
3. For the RG method, we apply to $\widehat{\theta}$ the change of scale which multiplies all loadings of indicators j belonging to the latent variable i by $\sqrt{\widehat{\phi}_{i, i, g_0}}$, latent variable i's standard deviation in the Reference Group g_0, while dividing the latent covariances $\widehat{\phi}_{i_1, i_2, g}$ between the latent variables i_1 and i_2 in Group g by $\sqrt{\widehat{\phi}_{i_1, i_1, g_0}} \cdot \sqrt{\widehat{\phi}_{i_2, i_2, g_0}}$. This transformation results in a new parameter for which the latent variable's variance in Group g_0 equals 1 (due to $\widetilde{\phi}^S_{i, i, g_0} = \frac{\widehat{\phi}_{i, i, g_0}}{\sqrt{\widehat{\phi}_{i, i, g_0}} \cdot \sqrt{\widehat{\phi}_{i, i, g_0}}} = 1$) and all indicators supposed to be invariant across groups stay invariant (as the already invariant loadings of $\widehat{\theta}$ are divided by a quantity which is also invariant across groups). Thus, the transformed

parameter fulfills the constraints associated with RG scaling as well as those of the partial metric MI model.

Corollary 6: If a partial metric MI model is estimated by minimizing a discrepancy function, then its χ^2, p value, RMSEA, and other fit measures do not depend on the particular method used for scaling the model, as long as the scaling method is applied in only one group.

This immediately follows from Proposition 2, as all these quantities are calculated using the partial metric MI model's likelihood value, which does not depend on the particular method used for scaling the model.

Corollary 7: The results of the χ^2-difference test which compares a partial metric MI model to the configural MI model do not depend on the scaling methods used for estimating the configural and partial metric MI model, as long as the scaling method used for the partial metric MI model is applied in only one group. This in particular holds for the χ^2-difference statistics, the p value, ΔRMSEA, and differences of other fit measures.

This immediately follows from the preceding corollary and the well-known fact that the corresponding quantities for the configural model do not depend on the particular method used for scaling the configural model, either.

Conclusions, Remarks, and Answers

In this paper, our goal was to clarify the impact of the various scaling methods on the estimation results for metric measurement invariance models. To this end, we addressed both full and partial metric MI models, and the results were laid out by means of worked examples as well as theoretical results with formal proofs. A first important insight of the paper is that scaling restrictions for metric MI models must be placed in one group only, which is of particular importance for partial metric MI models if the FM or EC methods are used and non-invariant loadings are involved. If the scaling restrictions are set in one group only, key quantities like optimal discrepancy values, χ^2, RMSEA, and other fit measures do neither depend on the choice of the scaling method in general nor on the choice of the RI in case of using the FM method. Thus, all the concerns about the choice of RIs mentioned in the literature can be put aside, as long as scaling is done properly: in particular, it does not matter whether a chosen RI is truly invariant or not. We are now ready to answer the first two questions: All scaling methods provide the same results in sense of the numerical value of the discrepancy function (and therefore all other fit indices) but the scaling must obey the rule, *Apply the scaling restriction in one group only, the invariance conditions do the rest!* The inconsistencies found in the literature regarding the RI choice may be a result of the lacking distinction between

dgp, scaled, and estimated model, which we introduced in the multiple-group context. Following this distinction, it is obvious that a non-invariant indicator in the dgp (in other words: a truly non-invariant indicator) does not affect the estimation of the scaled model. Irrespective of the nature of the dgp, the estimated model, which is stipulated by the researcher, provides the same value of the discrepancy function regardless of the scaling. As an extreme example, the dgp could be completely different from a confirmatory factor analysis model, but an estimated confirmatory factor analysis model would still provide the same results regardless of the applied scaling method. However, in this paper, we adopted a realistic philosophy of science and made the more or less implicit assumption that the dgp matches the estimated model structurally (cf., Klopp & Klößner, 2021, pp. 191, 202).

For partial metric MI models, the issues discussed in the literature regarding scaling and in particular RI selection originate from scaling restrictions being set in all groups, instead of in one group only. For instance, when using the FM method, one implicitly adds the RI to the set of indicators that are being tested empirically for loading invariance, even if one originally did not want to presume invariance of this indicator's loadings.[15] Thus, choosing different RIs leads to different sets of indicators whose loadings' invariance is under examination. Consequentially, the results differ, a phenomenon known as constraint interaction. These problems, however, can easily be avoided by setting scaling restrictions properly, i.e. by placing them in one group only. Apart from scaling in all groups instead of in only one, some of the concerns regarding the RI selection as well as concerns regarding other scaling methods, e.g., that the RG method implies an invariance assumption about the latent variances in the reference group (cf., Kline, 2016, p. 405), result from confounding the characteristics of the dgp with the scaling restrictions in the estimated model. As we have illustrated in Figure 1, these are distinctive entities.

One of the surprising results of this paper is that it is impossible to test the invariance or non-invariance of *one specific* loading by restricting only this loading to be invariant across groups, because in this case a partial metric MI model with the correct number of degrees of freedom is equivalent to the configural model. However, concerning the scaling of (partial) metric MI models, our research also showed that it is not necessary to choose an invariant RI, as long as the scaling is done in accordance with the rule introduced in this paper. This provides the answer to the third question: Also in partial metric MI models, all scaling methods provide the same results in sense of the numerical value of the discrepancy function (and therefore all other fit indices) but the scaling must obey the rule, *Apply the scaling restriction in one group only, the invariance conditions do*

15) The semTools package's help page for the function measEq.syntax correspondingly warns users that the RI's loadings can not be freed.

the rest! However, in partial metric MI settings, the researcher has to bear in mind that it is not possible to test the invariance of only one indicator.

With regard to the number of degrees of freedom for full and partial metric MI models, we provided formulas for calculating these easily. Given the findings of Schroeders and Gnambs (2018) regarding published papers with discrepancies with respect to the reported degrees of freedom, these formulas could prove very useful to applied researchers, when they try to determine the correct number of degrees of freedom for their models. Observing our rule, *Apply the scaling restriction in one group only, the invariance conditions do the rest!* in combination with the formulas for the degrees of freedom even provides a means for checking whether the scaling restrictions were set correctly.

The corollaries concerning the degrees of freedom also point at another issue. For the FM scaling method, Raykov et al. (2012, Appendix A) mentioned that the test of metric MI is incomplete because the MI model only tests the group equality of the scaled[16] subset of indicators but not the RI. As the test statistic does not depend on the scaling method, the RG and EC scaling methods also provide an incomplete test. However, the term incomplete as introduced by Raykov et al. (2012) in the context of the FM scaling method cannot be applied to the other two scaling methods. The corollaries provide firstly an explanation why the test is incomplete and secondly, explain what incomplete means in the context of the RG and EC. As can be seen in the corollaries about the degrees of freedom in the $(G-1)m$ term, one degree of freedom is lost due to the scaling restriction in the reference group regardless of the scaling method and this restriction is then propagated to the other groups via the invariance conditions. Thus, all scaling method have in common that they lose one degree of freedom due to the need to scale the metric MI model and that represents a meaning of the term incomplete which is common for all scaling methods. In addition, as all scaling methods are equivalent, the RI is not somehow excluded from invariance testing, as it is treated in the same way as the other indicators that are presumed to be invariant. And, finally, the underlying hypothesis being tested empirically does not depend on the scaling method a researcher decides to employ.

Concerning the fourth and last question, all scaling methods have the same mechanism: the scaling restriction is only applied in the reference group and propagated by means of the MI conditions to the other groups. Thus, there are again no differences between the various scaling method in this regards. In particular, this mechanism is the foundation for the, *Apply the scaling restriction in one group only, the invariance conditions do the rest!* rule that is of utmost importance in the scaling of metric MI models.

To sum up, there are no potential issues concerning the choice of the RI when using the FM scaling method and more generally, all scaling methods yield the same result. Consequently, any chapter concerning this topic can be deleted from the textbooks.

16) Raykov et al. (2012) use the term rescaled which in our framework refers to the scaled model.

Instead, a thorough explanation of the scaling methods and the rule, *Apply the scaling restriction in one group only, the invariance conditions do the rest!* should be included. In order to set the focus on the scaling issues of the metric model, we did not address the scaling of scalar measurement invariance models, which in contrast to metric MI models additionally incorporate a mean structure. We expect that future research will produce results similar to the current paper, probably by using techniques resembling the ones used in this paper.

At this point, we want to emphasize that our consideration refers in the first line to the discrepancy function and, and in turn, to the LR-test in the form of the χ^2-difference tests in the model testing sequence (see the Introduction section). The results should not be naively generalized to Wald or score tests, that are also sometimes used to investigate measurement invariance models. However, the developed scaling rule and the consideration about the degrees of freedom apply regardless of the respective statistical test.

We want to note that some parts of the results we presented are already present in the current literature. For instance, Wu and Estabrook (2016) provide a comprehensive account on the identification of confirmatory factor analysis models. However, they focus on the special case of models with ordinal indicator and scrutinize the identification of models with combinations of some invariance conditions. They also only consider the RG scaling method. In contrast, the approach in this article is to focus only on the scaling of the metric MI model considering the most common scaling methods. As mentioned above, we left the scaling restriction of the intercepts (or thresholds) behind.

Finally, we want to note that we also did not turn our attention to factors that potentially have effects on metric MI tests, e.g., the size of the manifest variables' residual variances. For instance, the example of Raykov et al. (2020) mentioned in the introduction, is the result of overlarge manifest residual error variances. As shown in Klopp and Klößner (2022), all possible scaling methods, i.e., all variants of the FM scaling as well as the RG and EC scaling methods, provide the same result test statistic. The authors also demonstrated by means of a Monte Carlo simulation, that lowering the size of the residual variances increased the chance to detect the violation of the metric MI condition. Another restriction was to focus on the multi-group context, ignoring longitudinal ones. However, all our examples and theoretical results can easily be translated to models with dependent variables, in particular, longitudinal models, where time essentially takes the role that groups take in this paper. In a nutshell, the scaling method does not affect results for longitudinal MI models of full or partial metric MI, as long as latent variables measured repeatedly over time are scaled by imposing a scaling restriction at only one point in time.

Funding: The authors have no funding to report.

Acknowledgments: The authors are grateful for valuable remarks from Sandra Baar, Martin Becker, Martin Klein, Mireille Soliman, and Holger Steinmetz (in alphabetical order).

Competing Interests: The authors have declared that no competing interests exist.

Data Availability: The R code to replicate the results of the current study, models and scenarios, as well as the supporting information, are freely available and can be found in the Supplementary Materials.

Supplementary Materials

The supplementary materials provided are the R code to replicate the results of the current study, models and scenarios, and supporting information (see Klopp & Klößner, 2022).

Index of Supplementary Materials

Klopp, E., & Klößner, S. (2022). *Scaling Metric Measurement Invariance Models* [R code, models and scenarios, supporting information]. OSF. https://osf.io/7qkth/

References

Bentler, P., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bollen, K. A. (1989). *Structural equations with latent variables.* Wiley. https://doi.org/10.1002/9781118619179

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*(1), 1–27. https://doi.org/10.1177/014920639902500101

Davidov, E., Muthen, B., & Schmidt, P. (2018). Measurement invariance in cross-national studies: Challenging traditional approaches and evaluating new ones. *Sociological Methods & Research, 47*(4), 631–636. https://doi.org/10.1177/0049124118789708

Greiff, S., & Scherer, R. (2018). Still comparing apples with oranges. *European Journal of Psychological Assessment, 34*(3), 141–144. https://doi.org/10.1027/1015-5759/a000487

Holland, P. W., & Thayer, D. T. (1988). Test validity. In H. Wainer & H. Braum, (Eds.), *Differential item performance and the Mantel-Haenszel procedure* (pp. 129–145). Hillsdale.

Johnson, E. C., Meade, A. W., & DuVernet, A. M. (2009). The role of referent indicators in tests of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(4), 642–657. https://doi.org/10.1080/10705510903206014

Jöreskog, K. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika, 43*, 443–477. https://doi.org/10.1007/bf02293808

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). *semTools: Useful tools for structural equation modeling* [R package version 0.5-5]. https://CRAN.R-project.org/package=semTools

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press. https://doi.org/10.4324/9781315749105

Klopp, E., & Klößner, S. (2021). The impact of scaling methods on the properties and interpretation of parameter estimates in structural equation models with latent variables. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(2), 182–206. https://doi.org/10.1080/10705511.2020.1796673

Klopp, E., & Klößner, S. (2022). The effects of scaling, manifest residual variances, and sample size on the χ^2-test statistic of the metric invariance model. In A. Hernández, & I. Tomás (Eds.), *Proceedings from the 9th European Congress of Methodology* (pp. 100–106). Universitat de Valencia. https://doi.org/10.7203/PUV-OA-438-5

Klößner, S., & Klopp, E. (2019). Explaining constraint interaction: How to interpret estimated model parameters under alternative scaling methods. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(1), 143–155. https://doi.org/10.1080/10705511.2018.1517356

Lee, J., Little, T. D., & Preacher, K. J. (2011). Methodological issues in using structural equation models for testing differential item functioning. In E. Davidov, P. Schmidt & J. Billet (Eds.), *Cross-cultural Analysis – Methods and Applications* (pp. 55–84). Psychology Press. https://doi.org/10.4324/9781315537078

Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & van de Schoot, R. (2023). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research, 110*, Article 102805. https://doi.org/10.1016/j.ssresearch.2022.102805

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in sem and macs models. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3

Millsap, R. (2011). *Statistical approaches to measurement invariance.* Routledge. https://doi.org/10.4324/9780203821961

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). CRC Press. https://doi.org/10.1201/b15851

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction.* Routledge. https://doi.org/10.4324/9781315871318

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

Raykov, T., Marcoulides, G. A., Harrison, M., & Zhang, M. (2020). On the dependability of a popular procedure for studying measurement invariance: A cause for concern? *Structural Equation Modeling: A Multidisciplinary Journal, 27*(4), 649–656. https://doi.org/10.1080/10705511.2019.1610409

Raykov, T., Marcoulides, G. A., & Li, C.-H. (2012). Measurement invariance for latent constructs in multiple populations: A critical view and refocus. *Educational and Psychological Measurement, 72*(6), 954–974. https://doi.org/10.1177/0013164412441607

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

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

Schroeders, U., & Gnambs, T. (2018). Degrees of freedom in multigroup confirmatory factor analysis – Are models of measurement invariance testing correctly specified? *European Journal of Psychological Assessment, 36*(1), 105–113. https://doi.org/10.1027/1015-5759/a000500

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78–90. https://doi.org/10.1086/209528

Steiger, J. H. (2002). When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. *Psychological Methods, 7*(2), 210–227. https://doi.org/10.1037/1082-989x.7.2.210

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

van de Schoot, R., Schmidt, P., & de Beuckelaer, A. (2015). *Measurement invariance*. Frontiers Media.

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

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer. http://www.stats.ox.ac.uk/pub/MASS4/

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika, 81*, 1014–1045. https://doi.org/10.1007/s11336-016-9506-0

Yoon, M., & Millsap, R. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 435–463. https://doi.org/10.1080/10705510701301677

Methodology is the official journal of the European Association of Methodology (EAM).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.