

Implementation Aspects in Invariance Alignment

Alexander Robitzsch ^{1,2} 

¹ IPN—Leibniz Institute for Science and Mathematics Education, Olshausenstraße 62, 24118 Kiel, Germany; robitzsch@leibniz-ipn.de

² Centre for International Student Assessment (ZIB), Olshausenstraße 62, 24118 Kiel, Germany

Abstract: In social sciences, multiple groups, such as countries, are frequently compared regarding a construct that is assessed using a number of items administered in a questionnaire. The corresponding scale is assessed with a unidimensional factor model involving a latent factor variable. To enable a comparison of the mean and standard deviation of the factor variable across groups, identification constraints on item intercepts and factor loadings must be imposed. Invariance alignment (IA) provides such a group comparison in the presence of partial invariance (i.e., a minority of item intercepts and factor loadings are allowed to differ across groups). IA is a linking procedure that separately fits a factor model in each group in the first step. In the second step, a linking of estimated item intercepts and factor loadings is conducted using a robust loss function $L_{0.5}$. The present article discusses implementation alternatives in IA. It compares the default $L_{0.5}$ loss function with L_p with other values of the power p between 0 and 1. Moreover, the nondifferentiable L_p loss functions are replaced with differentiable approximations in the estimation of IA that depend on a tuning parameter ε (such as, e.g., $\varepsilon = 0.01$). The consequences of choosing different values of ε are discussed. Moreover, this article proposes the L_0 loss function with a differentiable approximation for IA. Finally, it is demonstrated that the default linking function in IA introduces bias in estimated means and standard deviations if there is noninvariance in factor loadings. Therefore, an alternative linking function based on logarithmized factor loadings is examined for estimating factor means and standard deviations. The implementation alternatives are compared through three simulation studies. It turned out that the linking function for factor loadings in IA should be replaced by the alternative involving logarithmized factor loadings. Furthermore, the default $L_{0.5}$ loss function is inferior to the newly proposed L_0 loss function regarding the bias and root mean square error of factor means and standard deviations.



Citation: Robitzsch, A. Implementation Aspects in Invariance Alignment. *Stats* **2023**, *6*, 1160–1178. <https://doi.org/10.3390/stats6040073>

Academic Editor: Wei Zhu

Received: 17 September 2023

Revised: 22 October 2023

Accepted: 23 October 2023

Published: 25 October 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: confirmatory factor analysis; multiple groups; measurement invariance; invariance alignment; alignment optimization; robust loss function

1. Introduction

In the comparison of multiple groups in confirmatory factor analysis (CFA), some identifying assumptions have to be made. It is frequently assumed that item parameters are equal across groups, which is denoted as measurement invariance [1]. The invariance concept has been very prominent in psychology and the social sciences in general [2,3]. For example, in international large-scale assessment studies in education, like the programme for international student assessment (PISA), the necessity of invariance is strongly emphasized [4].

In violation of measurement invariance, the invariance alignment (IA) method [5,6] has been proposed to achieve approximate invariance [7]. That is, item parameters should be made as invariant as possible while allowing a few deviations from invariance. By doing so, group comparisons can be made more robust against violations of measurement invariance. Note that IA is also referred to as alignment optimization [8,9].

Although IA can be seen as a canonical method for handling measurement noninvariance in the social sciences [10,11], this method has not been thoroughly studied from a

statistical and conceptual point of view. There are a few simulation studies that investigate the behavior of the IA method. Except for a few studies [12,13], all simulation studies were carried out with the popular but commercial (and closed-source) Mplus software [14]. Previous simulation studies for unidimensional factor models investigated the case of continuous items [5,9,15–17], dichotomous items [18,19], and polytomous items [13,20]. The extension of IA to multidimensional factor models with continuous items was discussed in [21,22]. IA was studied in longitudinal measurement models in [23–26]. The IA method has been extended to exploratory structural equation models [22,27]. Moreover, the optimization function used in IA gave rise to extending it to a general framework used in the penalized maximum likelihood estimation of structural equation models [28].

IA has been applied in a wide area of disciplines. For example, IA has been utilized to compare European countries regarding attitudes towards migration [29,30]. Ref. [31] compares seven Latin American countries for the purpose in life test through IA. IA was applied to study bullying for children and adolescents [32,33]. Questionnaire data from the programme for international student assessment (PISA) study [34,35] and the trends in international mathematics and science study (TIMSS) [36] were used in IA applications. Furthermore, the IA method was utilized to investigate overexcitability [37], homophobia [38], distributed leadership [39], and gender role attitudes [40].

In this article, we focus on the implementation aspects of the IA method. The IA optimization function is nondifferentiable. Software implementations of the IA method rely on differentiable approximations that depend on a tuning parameter ε . The default value of this tuning parameter ε is critically examined in this paper. Furthermore, the originally proposed IA method utilizes the L_p loss function $\rho(x) = |x|^p$ for $p = 0.5$. This article investigates whether other choices than $p = 0.5$ result in improved estimation performance of the IA method. Moreover, the IA method uses a particular linking function for determining factor standard deviation based on a quantification of differences in residual item loadings. We show in this article that the performance of IA can be improved by relying on a different linking function that employs logarithmized item loadings to estimate factor standard deviations. Finally, the performance of IA is compared with a recently proposed differentiable approximation of the L_0 loss function. It turns out that this loss function performed comparably to default IA implementations, if not better, regarding the root mean square error of parameter estimates.

The rest of this article is organized as follows. IA estimation based on the robust L_p and L_0 loss functions is treated in Section 2. Section 3 discusses the standard error computation in IA. In Section 4, the research purpose of this article is outlined. Sections 5–7 contain three simulation studies, respectively, that thoroughly investigate the choices in IA implementation. Finally, the article closes with a discussion in Section 8.

2. Loss Functions in Invariance Alignment

In this section, the statistical background of IA is reviewed. In particular, the choice of different loss functions in IA is discussed.

Let X_{ig} denote item i ($i = 1, \dots, I$) in group g ($g = 1, \dots, G$). A unidimensional factor model [41] is defined as

$$X_{ig} = \nu_{ig} + \lambda_{ig}F_g + \epsilon_{ig}, \quad F_g \sim N(\mu_g, \sigma_g^2), \quad \epsilon_{ig} \sim N(0, \omega_{ig}), \quad (1)$$

where λ_{ig} is an item loading and ν_{ig} is an item intercept. Without loss of generality, item loadings can be assumed to be positive. The factor variables F_g and all residual variables ϵ_{ig} are independent and normally distributed. The factor variable F_g has a factor mean μ_g and a factor standard deviation σ_g .

It must be emphasized that the model parameters in (1) are not identified. An identified model is obtained by assuming a standardized latent variable F_g (i.e., with a mean of 0 and a standard deviation of 1):

$$X_{ig} = \nu_{ig,0} + \lambda_{ig,0}F_g + \epsilon_{ig}, \quad F_g \sim N(0, 1), \quad \epsilon_{ig} \sim N(0, \omega_{ig}). \quad (2)$$

The model parameters in (1) and (2) are related to each other by

$$\lambda_{ig,0} = \lambda_{ig} \sigma_g \text{ and } \nu_{ig,0} = \nu_{ig} + \lambda_{ig} \mu_g = \nu_{ig} + \frac{\lambda_{ig,0}}{\sigma_g} \mu_g. \tag{3}$$

A convenient property is measurement invariance [1,3], in which the same item loadings and item intercepts across groups can be assumed. That is, there exist item loadings λ_i such that $\lambda_i = \lambda_{ig}$ for all $g = 1, \dots, G$ and item intercepts ν_i such that $\nu_i = \nu_{ig}$ for $g = 1, \dots, G$ for all items $i = 1, \dots, I$. The absence of measurement invariance is also labeled as differential item functioning (DIF; [2,42]) in the literature. In the case of measurement invariance, Equation (3) can be rewritten as

$$\lambda_{ig,0} = \lambda_i \sigma_g \text{ and } \nu_{ig,0} = \nu_i + \frac{\lambda_{ig,0}}{\sigma_g} \mu_g. \tag{4}$$

The IA method of Asparouhov and Muthén [5,6] tackles situations under sparse violations of measurement invariance. In this case, a few item loadings or item intercepts are allowed to differ across groups, while the majority of items (approximately) fulfill the invariance assumption [43]. This situation is referred to as partial invariance [44].

In IA, the unidimensional factor model (2) is separately estimated for all groups in the first step. The estimated item parameters $\hat{\lambda}_{ig,0}$ and $\hat{\nu}_{ig,0}$ ($i = 1, \dots, I; g = 1, \dots, G$) are used as the input of the IA. By rewriting (3) and inserting the estimated item loadings and item intercepts, we obtain

$$\lambda_{ig} - \lambda_{ih} = \frac{\hat{\lambda}_{ig,0}}{\sigma_g} - \frac{\hat{\lambda}_{ih,0}}{\sigma_h} \text{ and } \nu_{ig} - \nu_{ih} = \hat{\nu}_{ig,0} - \hat{\nu}_{ih,0} - \frac{\hat{\lambda}_{ig,0}}{\sigma_g} \mu_g + \frac{\hat{\lambda}_{ih,0}}{\sigma_h} \mu_h. \tag{5}$$

These relations motivate the minimization of the following linking function in IA to determine group means $\mu = (\mu_1, \dots, \mu_G)$ and standard deviations $\sigma = (\sigma_1, \dots, \sigma_G)$:

$$H(\mu, \sigma) = \sum_{i=1}^I \sum_{g=1}^{G-1} \sum_{h=g+1}^G w_{i1,gh} \rho \left(\frac{\hat{\lambda}_{ig,0}}{\sigma_g} - \frac{\hat{\lambda}_{ih,0}}{\sigma_h} \right) + \sum_{i=1}^I \sum_{g=1}^{G-1} \sum_{h=g+1}^G w_{i2,gh} \rho \left(\hat{\nu}_{ig,0} - \hat{\nu}_{ih,0} - \frac{\hat{\lambda}_{ig,0}}{\sigma_g} \mu_g + \frac{\hat{\lambda}_{ih,0}}{\sigma_h} \mu_h \right), \tag{6}$$

where the weights $w_{i1,gh}$ and $w_{i2,gh}$ are known, and ρ is a loss function. Asparouhov and Muthén [5] proposed $w_{i1,gh} = w_{i2,gh} = \sqrt{n_g n_h}$ and $\rho(x) = \sqrt{|x|}$, where n_g denotes the sample size of group g . In the minimization of (6), additional identification constraints must be imposed. In this article, we fix the moments in the first group; that is, we set $\mu_1 = 0$ and $\sigma_1 = 1$.

In the rest of the article, we ignore the weights in (6) for the following reasons. First, it simplifies mathematical notation. Second, it is not obvious why one should choose weights related to the sample sizes of the groups. We think that model deviations should be equally weighted across groups.

Note that the optimization function H of IA, defined in (6), can be rewritten as

$$H(\mu, \sigma) = H_1(\sigma) + H_2(\mu, \sigma), \text{ where} \tag{7}$$

$$H_1(\sigma) = \sum_{i=1}^I \sum_{g=1}^{G-1} \sum_{h=g+1}^G \rho \left(\frac{\hat{\lambda}_{ig,0}}{\sigma_g} - \frac{\hat{\lambda}_{ih,0}}{\sigma_h} \right) \text{ and } H_2(\mu, \sigma) = \sum_{i=1}^I \sum_{g=1}^{G-1} \sum_{h=g+1}^G \rho \left(\hat{\nu}_{ig,0} - \hat{\nu}_{ih,0} - \frac{\hat{\lambda}_{ig,0}}{\sigma_g} \mu_g + \frac{\hat{\lambda}_{ih,0}}{\sigma_h} \mu_h \right). \tag{8}$$

It has been shown that the simultaneous minimization of H with respect to μ and σ can be viewed as a two-step minimization problem [45]. In more detail, a vector of estimated factor standard deviations $\hat{\sigma}$ is obtained by minimizing $H_1(\sigma)$ in the first step. In the second step, a vector of estimated factor means $\hat{\mu}$ is obtained by minimizing $H_2(\mu, \hat{\sigma})$ with respect to μ .

Equation (3) can be rewritten as

$$\log \lambda_{ig,0} = \log \lambda_{ig} + \log \sigma_g \text{ and } \nu_{ig,0} = \nu_{ig} + \frac{\lambda_{ig,0}}{\sigma_g} \mu_g. \quad (9)$$

This motivates using an alternative optimization function H_1^* for determining standard deviations that employs logarithmized item loadings (see [45,46])

$$H_1^*(\sigma^*) = \sum_{i=1}^I \sum_{g=1}^{G-1} \sum_{h=g+1}^G \rho \left(\log \hat{\lambda}_{ig,0} - \log \hat{\lambda}_{ih,0} - \sigma_g^* + \sigma_h^* \right), \quad (10)$$

where $\sigma_g^* = \log \sigma_g$ for $g = 1, \dots, G$. Due to the identification constraint, we fix $\sigma_1^* = 0$ (i.e., $\sigma_1 = \exp(\sigma_1^*) = 1$). By minimizing H_1^* , a vector of standard deviations $\hat{\sigma}^*$ on the logarithm metric is obtained; that is, $\hat{\sigma}^* = (\hat{\sigma}_1^*, \dots, \hat{\sigma}_G^*)$. The vector of estimated standard deviations $\hat{\sigma}$ can be obtained by exponentiating all entries in $\hat{\sigma}^*$. The vector of estimated factor means $\hat{\mu}$ can again be obtained by minimizing $H_2(\mu, \hat{\sigma})$.

Hence, there are two estimation options for IA. The original approach of [5] minimizes H_1 and is referred to as the “NOL” method (i.e., no logarithm for item loadings). The second approach obtains factor standard deviations by minimizing H_1^* and is referred to as the “LOG” method (i.e., taking the logarithmized item loadings for defining deviations).

As mentioned above, IA uses the loss function $\rho(x) = \sqrt{|x|} = |x|^{0.5}$ as the default in the Mplus software package [14]. However, the loss function $\rho(x) = |x|^{0.25}$ is also available in Mplus [14]. The more general L_p loss function $\rho(x) = |x|^p$ for $p > 0$ was studied for IA in [12,45]. It has been shown that values of the power p smaller than 0.5 can be advantageous in some situations [12]. An interesting case is the limiting case in which p tends to zero. Effectively, $p = 0$ counts the number of parameter deviations that differ from zero [47,48], resulting in the L_0 loss function. In the practical minimization of H involved in IA, the nondifferentiable L_p loss function $\rho(x) = |x|^p$ (for $0 < p \leq 1$) is replaced by a differentiable approximation ρ_D (see [5,45])

$$\rho_D(x) = (x^2 + \varepsilon)^{p/2}, \quad (11)$$

where $\varepsilon > 0$ is a tuning parameter that controls the approximation error of ρ_D for ρ . The approximation error becomes smaller with ε values close to zero. However, the minimization of H in IA gets more difficult when choosing too-small values of ε . Practical experience led to proposals $\varepsilon = 0.01$ [5] or $\varepsilon = 0.001$ [45]. The choice $\varepsilon = 0.01$ is the default in Mplus (see [12]). It is also tempting to consider the L_0 loss function $\rho(x) = |x|^0$ that takes values of 1 if $x \neq 0$ and 0 for $x = 0$. However, the differentiable approximation ρ_D in (11) performs poorly for p values close to 0 because the minimization in H becomes very difficult. O'Neill and Burke [49] proposed, in a recent work related to regularized estimation, the following differentiable approximation ρ_D of the L_0 loss function:

$$\rho_D(x) = \frac{x^2}{x^2 + \varepsilon}, \quad (12)$$

where $\varepsilon > 0$ is again a tuning parameter that controls the approximation error and estimation stability of the differentiable approximation. This approximation (12) of the L_0 loss function has not yet been investigated in IA. As IA is particularly suited to the sparse deviations in model parameters, the L_0 loss function should theoretically fit typical data-generating models frequently utilized in simulation studies that investigate the performance of IA.

In our experience, in the case of small ϵ values, the optimization of the alignment function is very sensitive to starting values. Asparouhov and Muthén [5] remark that the linking function in invariance alignment is prone to multiple local minima. Moreover, they mention that these local minima often yield values of the linking function that are only slightly different from values at the global minimum. In the estimation of the IA approach, it is advised to choose a sequence of decreasing values of ϵ in the optimization, each using the previous solution as initial values (see [50] for a similar approach). This choice guarantees better suitable starting values and a more stable estimation of IA.

3. Standard Errors in Invariance Alignment

In this section, the estimation of standard errors for the IA approach is described. IA minimizes the linking function $H(\mu, \sigma)$ with respect to μ and σ . Let $\theta = (\mu, \sigma)$ be the parameter vector of interest. The estimated item loadings and item intercepts across all groups are collected in a vector $\hat{\xi} = (\hat{\xi}_1, \dots, \hat{\xi}_G)$, where $\hat{\xi}_g$ contains the group-specific model parameter estimates from a unidimensional factor model in group $g = 1, \dots, G$. As a maximum likelihood estimate, $\hat{\xi}_g$ is approximately multivariate normally distributed. Because of the independence of subjects across groups, a multivariate normal distribution of the input parameter $\hat{\xi}$ in IA is obtained as

$$\hat{\xi} - \xi_0 \sim \text{MVN}(\mathbf{0}, V_{\xi}), \tag{13}$$

where V_{ξ} is a block-diagonal covariance matrix, and ξ_0 is a population parameter.

The distribution of the IA estimates $\hat{\theta}$ is now derived using the delta method in M-estimation theory [51] by relying on the implicit function theorem [5]. We assume differentiability of the optimization function because the nondifferentiable loss function ρ in IA is replaced by a differentiable approximation $\rho_{\mathcal{D}}$. The IA approach minimizes $H(\theta, \hat{\xi})$, where we now highlight the dependency of the input parameters $\hat{\xi}$. A parameter estimate $\hat{\theta}$ is obtained by taking the partial derivative of H with respect to θ (i.e., H_{θ}) and solving the nonlinear equation such that

$$H_{\theta}(\hat{\theta}, \hat{\xi}) = \mathbf{0}. \tag{14}$$

Note that there exists a population parameter θ_0 such that

$$H_{\theta}(\theta_0, \xi_0) = \mathbf{0}. \tag{15}$$

Now, a Taylor expansion of H_{θ} around (θ_0, ξ_0) can be carried out. Denote with $H_{\theta\theta}$ and $H_{\theta\xi}$ the matrices of second-order partial derivatives of H_{θ} with respect to θ and ξ , respectively. The Taylor expansion can be written as

$$H_{\theta}(\hat{\theta}, \hat{\xi}) = H_{\theta}(\theta_0, \xi_0) + H_{\theta\theta}(\theta_0, \xi_0)(\hat{\theta} - \theta_0) + H_{\theta\xi}(\theta_0, \xi_0)(\hat{\xi} - \xi_0) = \mathbf{0}. \tag{16}$$

By solving (16) for $\hat{\theta}$, we have the approximation

$$\hat{\theta} - \theta_0 = -H_{\theta\theta}(\theta_0, \xi_0)^{-1} H_{\theta\xi}(\theta_0, \xi_0)(\hat{\xi} - \xi_0). \tag{17}$$

By defining $\hat{A} = -H_{\theta\theta}(\hat{\theta}, \hat{\xi})^{-1} H_{\theta\xi}(\hat{\theta}, \hat{\xi})$ when substituting θ_0 and ξ_0 with $\hat{\theta}$ and $\hat{\xi}$, respectively, we have, by using the multivariate delta method [51],

$$\text{Var}(\hat{\theta}) = \hat{A} V_{\xi} \hat{A}^{\top}. \tag{18}$$

Standard errors for elements in $\hat{\theta}$ can be obtained by taking the square root of diagonal elements of $\text{Var}(\hat{\theta})$ computed from (18).

4. Purpose

In this article, several implementation aspects for IA should be examined. First, it should be investigated whether the choice of the power p impacts the performance of the IA estimates. In particular, it is interesting whether there are better alternatives to the default choice $p = 0.5$. Second, the deliberate choice of the tuning parameter ε is studied. It is interesting to researchers and developers of statistical software which ε should be selected as a default in order to produce the most reliable parameter estimates. It is expected that a larger value of ε results in a less precise differentiable approximation and more bias for the IA estimates compared with that expected with smaller values of ε . Third, the L_p loss function for $0 < p \leq 1$ should be compared with the newly proposed differentiable approximation of the L_0 loss function by O'Neill and Burke. It would be interesting to see whether the L_0 loss function could also be beneficial in invariance alignment. Fourth, the two different choices of the estimation of standard deviations σ (methods "NOL" and "LOG"; see Section 2) are compared. The NOL method uses deviations of $\hat{\lambda}_{ig,0}/\sigma_g$ and $\hat{\lambda}_{ih,0}/\sigma_h$ with respect to the loss function ρ , while the LOG method utilizes deviations $\log \hat{\lambda}_{ig,0} - \log \sigma_g$ and $\log \hat{\lambda}_{ih,0} - \log \sigma_h$. Fifth, and finally, the quality of standard errors in terms of coverage is studied for the different IA estimation approaches. These research questions were answered by means of three simulation studies that are described in the next three sections.

5. Simulation Study 1: Bias and RMSE in a Three-Group Example

In Simulation Study 1, IA is studied in a case with noninvariant item intercepts and in a case with noninvariant item loadings and noninvariant item intercepts. This study focuses on bias and root mean square error (RMSE).

5.1. Method

The data-generating models (DGMs) in the simulation study mimicked the DGM used in [5]. The data were simulated from a one-dimensional factor model involving five items (i.e., $I = 5$) and three groups (i.e., $G = 3$). The factor variable was normally distributed with group means 0, 0.3, and 0.8, and the group variances were 1, 1.5, and 1.2, respectively. All measurement error variances were set to one in all groups and were uncorrelated with each other. The factor variable and residual variables were normally distributed.

Two DGMs were simulated that refer to a violation of measurement invariance. Group-specific item parameters that are noninvariant are referred to as DIF effects [2].

In the first DGM (i.e., DGM1), only item intercepts were noninvariant. All item loadings were set to one, and only a subset of group-specific item intercepts were simulated differently from zero. Hence, data were simulated assuming partial invariance. In the first group, the fourth item intercept was 0.5. In the second group, the first item was -0.5 , while the second item had an intercept of -0.5 in the third group.

In the second DGM (i.e., DGM2), item intercepts and item loadings were invariant. The same item intercepts as in DGM1 were used. Three group-specific item loadings were different from one. The item loading of the third item in the first group and the item loadings of the fifth item in the third group were 2.013. The second item in the second group had an item loading of 0.497.

The sample size per group was chosen as $N = 250$, $N = 500$, $N = 1000$, or $N = 2000$. IA was estimated using the NOL method (based on the H_1 function defined in (8)) and the LOG method (based on the H_1^* function defined in (10)). The L_p loss function was employed using the powers $p = 0.5$, $p = 0.25$, and $p = 0.1$ and the differentiable approximation defined in (11). The L_0 loss function employed differentiable approximation (12). We did not consider power values $p = 2$ or $p = 1$ because they have been shown to result in severely biased estimates in the situation of partial invariance [12,45]. The reason is that noninvariant item intercepts (i.e., model errors) indicate a misspecified model. This kind of misspecification biases factor means if the L_p loss function is with $p = 2$ because all item intercepts contribute to the estimation of factor means. The situation is known from robust

statistics where the unweighted mean is not robust to outlying observations. Moreover, the L_p loss function with $p = 1$ does not fully remove bias because it treats model errors (i.e., outlying observations) symmetrically. In contrast, the L_p loss function with $p < 1$ is more robust for asymmetrically distributed model errors.

All estimation methods were applied with the tuning parameters $\varepsilon = 0.1$, $\varepsilon = 0.01$, $\varepsilon = 0.001$, and $\varepsilon = 0.0001$. The choice $\varepsilon = 0.1$ led to substantially biased parameter estimates, while the IA estimates based on $\varepsilon = 0.0001$ had large variances. Hence, we only report findings for the tuning parameters $\varepsilon = 0.01$ and $\varepsilon = 0.001$.

In total, $R = 1000$ replications were conducted for each cell of the simulation study. Bias, standard deviation (SD), RMSE, and relative RMSE were computed to assess the performance of the different estimators. Let $\hat{\theta}_{j,r}$ be a model parameter estimate in replication $r = 1, \dots, R$ for the parameter θ_j . The bias of the estimator $\hat{\theta}_j$ was estimated with

$$\text{Bias}(\hat{\theta}_j) = \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_{j,r} - \theta_j), \quad (19)$$

where θ_j denotes the true parameter value. The SD of an estimator $\hat{\theta}_j$ was calculated as

$$\text{SD}(\hat{\theta}_j) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_{j,r} - \bar{\theta}_{j,\bullet})^2}, \text{ where } \bar{\theta}_{j,\bullet} = \frac{1}{R} \sum_{r=1}^R \hat{\theta}_{j,r}. \quad (20)$$

The RMSE of an estimator $\hat{\theta}_j$ was estimated with

$$\text{RMSE}(\hat{\theta}_j) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_{j,r} - \theta_j)^2}. \quad (21)$$

A relative RMSE can be defined by dividing the RMSE of an estimator by the RMSE of a chosen reference model. In Simulation Study 1 (and in Simulation Study 3), the Mplus default with the NOL method, $p = 0.5$, and $\varepsilon = 0.01$ is used as the reference model. To more easily grasp differences in the relative RMSE, the values were multiplied by 100. This quantity can then easily be converted into a percentage gain of a particular estimator compared with a reference model.

The entire simulation study was carried out in the R [52] software. IA was performed with the `sirt::invariance.alignment()` (see also [12,53,54]) function in the R package `sirt` (Version 4.0-19; [55]). Information about model specification can be found in the material located at <https://osf.io/7kwqh/> (accessed on 17 September 2023).

5.2. Results

Table 1 reports the bias, SD, and relative RMSE of the factor mean μ_2 and the factor SD σ_2 of the second group in the DGM of noninvariant item intercepts (i.e., DGM1). It can be seen that the L_p loss function with $p = 0.5$, $p = 0.25$, and $p = 0.1$ showed some bias for μ_2 . Importantly, the bias was more substantial when using $\varepsilon = 0.01$ instead of $\varepsilon = 0.001$. Furthermore, the extent of the bias in the estimated factor mean μ_2 decreased with increasing sample size. Notably, the pattern of the bias was similar for the NOL and LOG methods. Interestingly, the L_0 loss function (i.e., $p = 0$) outperformed the other specifications regarding bias. While $\varepsilon = 0.001$ would be preferable for $p = 0.5$, $p = 0.25$, and $p = 0.1$, for $p = 0$, the tuning parameter choice $\varepsilon = 0.01$ would be preferred due to a smaller SD of the estimate. Notably, $p = 0$ had a slightly increased SD compared with $p = 0.5$ for the sample size $N = 250$. However, this effect decreased with larger sample sizes. Moreover, it can be seen that the Mplus default $p = 0.5$ and $\varepsilon = 0.01$ could be improved in terms of relative RMSE by using $p = 0.5$ and $\varepsilon = 0.001$ for DGM1. Notably, the relative performance gains are more important in larger sample sizes. Additional smaller gains can be obtained by switching to $p = 0$ and $\varepsilon = 0.01$.

The factor SD σ_2 was almost unbiasedly estimated in the condition of only noninvariant item intercepts. In this situation, the choice $\varepsilon = 0.01$ can be defended over $\varepsilon = 0.001$ for the Mplus default. The L_0 loss function could not outperform p values different from zero regarding the SD.

Table 1. Simulation Study 1: bias, standard deviation (SD), and relative root mean square error (RMSE) of the factor mean μ_2 and the factor standard deviation σ_2 as a function of sample size N for different estimation methods in the case of noninvariant item intercepts (DGM1).

Meth	p, ε	Bias				SD				RMSE			
		N				N				N			
		250	500	1000	2000	250	500	1000	2000	250	500	1000	2000
<i>Factor Mean μ_2</i>													
NOL	0.5, 0.01 [†]	-0.067	-0.051	-0.041	-0.037	0.118	0.082	0.058	0.040	100 [†]	100 [†]	100 [†]	100 [†]
	0.5, 0.001	-0.043	-0.027	-0.016	-0.012	0.124	0.086	0.059	0.041	96.8	93.2	86.5	77.5
	0.25, 0.01	-0.050	-0.036	-0.027	-0.024	0.120	0.083	0.058	0.040	95.9	92.9	90.3	86.0
	0.25, 0.001	-0.031	-0.017	-0.009	-0.006	0.127	0.087	0.059	0.041	96.6	91.7	85.1	75.6
	0.1, 0.01	-0.042	-0.029	-0.021	-0.019	0.121	0.083	0.058	0.040	94.5	90.4	87.1	81.3
	0.1, 0.001	-0.026	-0.013	-0.007	-0.004	0.128	0.087	0.060	0.041	96.6	90.4	84.9	75.4
	0, 0.01	-0.007	-0.001	0.000	0.000	0.128	0.084	0.058	0.040	94.8	86.8	82.8	73.6
	0, 0.001	-0.006	0.000	0.001	0.000	0.133	0.089	0.063	0.043	98.4	92.3	89.6	79.1
LOG	0.5, 0.01	-0.068	-0.052	-0.041	-0.037	0.117	0.082	0.057	0.040	100.1	100.2	100.2	100.2
	0.5, 0.001	-0.045	-0.028	-0.016	-0.012	0.123	0.086	0.059	0.041	96.7	93.2	86.5	77.6
	0.25, 0.01	-0.051	-0.036	-0.028	-0.024	0.119	0.082	0.057	0.040	95.9	93.0	90.4	86.1
	0.25, 0.001	-0.033	-0.018	-0.009	-0.007	0.126	0.087	0.059	0.041	96.3	91.6	85.0	75.6
	0.1, 0.01	-0.044	-0.030	-0.022	-0.019	0.120	0.082	0.058	0.040	94.5	90.5	87.2	81.4
	0.1, 0.001	-0.028	-0.014	-0.007	-0.005	0.127	0.086	0.059	0.041	96.3	90.2	84.8	75.4
	0, 0.01	-0.009	-0.002	0.000	0.000	0.128	0.084	0.058	0.040	94.4	86.5	82.7	73.6
	0, 0.001	-0.008	-0.001	0.000	-0.001	0.132	0.089	0.063	0.043	97.8	92.0	89.5	79.0
<i>Factor SD σ_2</i>													
NOL	0.5, 0.01 [†]	0.011	0.006	0.001	0.001	0.096	0.067	0.045	0.033	100 [†]	100 [†]	100 [†]	100 [†]
	0.5, 0.001	0.013	0.006	0.001	0.001	0.101	0.070	0.047	0.033	105.5	104.1	103.2	102.2
	0.25, 0.01	0.011	0.006	0.001	0.001	0.097	0.067	0.046	0.033	100.6	100.3	100.2	100.1
	0.25, 0.001	0.014	0.006	0.001	0.000	0.104	0.071	0.048	0.034	108.2	106.2	104.8	103.0
	0.1, 0.01	0.012	0.006	0.001	0.001	0.097	0.067	0.046	0.033	101.1	100.5	100.3	100.1
	0.1, 0.001	0.013	0.006	0.002	0.000	0.105	0.072	0.048	0.034	109.2	107.2	105.9	103.6
	0, 0.01	0.013	0.006	0.001	0.001	0.105	0.070	0.047	0.033	109.1	104.7	102.6	101.1
	0, 0.001	0.015	0.006	0.002	0.000	0.114	0.077	0.054	0.037	119.1	115.3	118.5	112.3
LOG	0.5, 0.01	0.004	0.003	-0.001	0.000	0.096	0.067	0.046	0.033	99.1	99.5	100.1	100.0
	0.5, 0.001	0.004	0.002	-0.001	0.000	0.101	0.069	0.047	0.033	104.2	102.9	103.1	102.2
	0.25, 0.01	0.004	0.003	-0.001	0.000	0.096	0.067	0.046	0.033	99.6	99.8	100.2	100.1
	0.25, 0.001	0.004	0.002	-0.001	0.000	0.103	0.070	0.048	0.034	106.8	104.6	104.6	103.1
	0.1, 0.01	0.004	0.003	-0.001	0.000	0.097	0.067	0.046	0.033	100.1	100.0	100.3	100.1
	0.1, 0.001	0.004	0.002	0.000	0.000	0.104	0.071	0.048	0.034	107.9	105.9	105.6	103.7
	0, 0.01	0.004	0.002	-0.001	0.000	0.105	0.070	0.047	0.033	108.7	104.0	102.6	101.1
	0, 0.001	0.004	0.001	0.000	-0.001	0.113	0.077	0.053	0.037	117.2	114.4	117.1	112.5

Note. Meth = estimation method; NOL = no logarithmized linking function for item loadings using H_1 in (8); LOG = logarithmized linking function for item loadings using H_1^* in (10); p = power used in the L_p loss function ρ ; ε = tuning parameter used in the differentiable approximation of the L_p loss function ρ ; absolute biases larger than 0.03 are shown with a gray background. [†] The Mplus defaults $p = 0.5$, $\varepsilon = 0.01$, and NOL are the reference methods in the computation of the relative RMSE. Relative RMSE values smaller than 95.0 are printed in bold font.

Table 2 displays bias, SD, and relative RMSE for the factor mean and the factor SD of the second group in the condition of noninvariant item loadings and noninvariant item intercepts. In this situation, the SD estimate was biased, particularly for the smallest sample size $N = 250$. Importantly, the bias was significantly reduced when using the LOG method instead of the NOL method. Overall, the L_0 loss function with $\varepsilon = 0.01$ was the preferred

method regarding the bias and relative RMSE for sample sizes of at least 500. Surprisingly, no bias occurred for the estimated factor mean in DGM2. However, this seems to be a coincidence of different defining factors for the bias. Table A1 in Appendix A reveals that the factor mean of the third group in DGM2 also provided biased estimates. Again, in this case, $p = 0$ resolved the issue, in particular for larger sample sizes.

Table 2. Simulation Study 1: bias, standard deviation (SD), and relative root mean square error (RMSE) of the factor mean μ_2 and the factor standard deviation σ_2 as a function of sample size N for different estimation methods in the case of noninvariant item intercepts and noninvariant item loadings (DGM2).

Meth	p, ϵ	Bias				SD				RMSE			
		N				N				N			
		250	500	1000	2000	250	500	1000	2000	250	500	1000	2000
<i>Factor Mean μ_2</i>													
NOL	0.5, 0.01 [†]	-0.016	-0.012	-0.008	-0.006	0.127	0.085	0.059	0.041	100 [†]	100 [†]	100 [†]	100 [†]
	0.5, 0.001	-0.008	-0.008	-0.004	-0.002	0.130	0.086	0.057	0.039	101.7	99.9	96.2	96.2
	0.25, 0.01	-0.007	-0.006	-0.004	-0.003	0.128	0.084	0.058	0.040	99.8	97.7	97.2	96.9
	0.25, 0.001	0.000	-0.003	-0.002	-0.001	0.132	0.086	0.058	0.039	102.5	100.3	96.7	95.8
	0.1, 0.01	-0.003	-0.004	-0.003	-0.002	0.129	0.084	0.057	0.039	100.2	97.0	96.2	95.8
	0.1, 0.001	0.003	-0.002	-0.001	0.000	0.133	0.086	0.058	0.039	103.8	99.9	96.8	95.7
	0, 0.01	0.015	0.003	0.001	0.001	0.136	0.084	0.056	0.038	106.5	97.2	94.5	92.7
	0, 0.001	0.016	0.004	0.001	0.001	0.140	0.088	0.060	0.042	109.5	102.2	100.3	102.0
LOG	0.5, 0.01	-0.031	-0.023	-0.017	-0.014	0.120	0.082	0.057	0.039	96.7	98.5	99.7	101.6
	0.5, 0.001	-0.021	-0.014	-0.007	-0.004	0.123	0.084	0.057	0.039	97.4	98.1	95.6	95.8
	0.25, 0.01	-0.022	-0.015	-0.011	-0.008	0.121	0.081	0.056	0.039	95.5	95.3	96.1	96.7
	0.25, 0.001	-0.014	-0.008	-0.004	-0.002	0.123	0.084	0.057	0.039	96.8	97.5	96.2	95.4
	0.1, 0.01	-0.017	-0.012	-0.008	-0.006	0.121	0.080	0.056	0.039	95.3	94.3	95.0	95.2
	0.1, 0.001	-0.011	-0.007	-0.003	-0.001	0.124	0.083	0.057	0.039	97.0	97.1	96.1	95.4
	0, 0.01	-0.001	-0.001	0.000	0.001	0.126	0.080	0.056	0.038	98.0	93.1	93.1	92.5
	0, 0.001	0.000	0.000	0.000	0.001	0.129	0.085	0.059	0.042	100.4	98.0	99.1	102.1
<i>Factor SD σ_2</i>													
NOL	0.5, 0.01 [†]	0.183	0.135	0.108	0.093	0.134	0.094	0.061	0.041	100 [†]	100 [†]	100 [†]	100 [†]
	0.5, 0.001	0.142	0.072	0.042	0.028	0.161	0.098	0.059	0.038	94.6	73.7	58.2	46.0
	0.25, 0.01	0.163	0.105	0.077	0.064	0.146	0.097	0.060	0.040	96.7	86.8	78.7	73.7
	0.25, 0.001	0.132	0.056	0.028	0.017	0.175	0.104	0.058	0.037	96.5	72.2	51.8	40.4
	0.1, 0.01	0.153	0.091	0.064	0.051	0.152	0.098	0.059	0.039	95.5	81.3	70.1	63.3
	0.1, 0.001	0.127	0.050	0.023	0.013	0.179	0.106	0.058	0.037	96.9	71.2	50.5	39.0
	0, 0.01	0.114	0.031	0.010	0.005	0.207	0.114	0.058	0.035	104.4	71.8	47.7	35.1
	0, 0.001	0.113	0.028	0.007	0.003	0.214	0.120	0.064	0.040	106.8	74.9	51.6	39.4
LOG	0.5, 0.01	0.108	0.081	0.066	0.058	0.117	0.081	0.054	0.037	70.3	69.4	68.7	67.5
	0.5, 0.001	0.078	0.044	0.027	0.019	0.128	0.084	0.054	0.037	66.2	57.8	49.1	40.7
	0.25, 0.01	0.094	0.064	0.049	0.041	0.121	0.081	0.054	0.037	67.6	62.7	58.5	54.4
	0.25, 0.001	0.069	0.034	0.018	0.012	0.133	0.086	0.054	0.037	66.1	56.4	45.9	38.0
	0.1, 0.01	0.086	0.055	0.041	0.034	0.124	0.081	0.053	0.036	66.7	59.8	54.2	49.1
	0.1, 0.001	0.065	0.029	0.014	0.009	0.135	0.087	0.054	0.037	66.3	55.9	45.1	37.2
	0, 0.01	0.047	0.013	0.004	0.003	0.147	0.083	0.051	0.035	68.3	51.4	41.2	34.6
	0, 0.001	0.046	0.010	0.003	0.002	0.152	0.090	0.057	0.040	70.1	54.9	45.9	39.2

Note. Meth = estimation method; NOL = no logarithmized linking function for item loadings using H_1 in (8); LOG = logarithmized linking function for item loadings using H_1^* in (10); p = power used in the L_p loss function; ϵ = tuning parameter used in the differentiable approximation of the L_p loss function; absolute biases larger than 0.03 are shown with a gray background. [†] The Mplus defaults $p = 0.5$, $\epsilon = 0.01$, and NOL are the reference methods in the computation of the relative RMSE. Relative RMSE values smaller than 95.0 are printed in bold font.

As a preliminary conclusion from Simulation Study 1, one could argue that the case of noninvariant item loadings can induce bias in parameter estimates in default implementations of IA. The bias can be reduced by using the LOG method instead of the default

NOL method. There is a tendency that the choice $\epsilon = 0.001$ outperformed $\epsilon = 0.01$ for $p = 0.5$. Finally, the L_0 loss function (i.e., $p = 0$) had satisfactory performance for $\epsilon = 0.01$. However, it came at the price of increased variance in smaller sample sizes.

6. Simulation Study 2: Coverage Rates in a Three-Group Example

The second example, Simulation Study 2, investigated the assessment of coverage rates for the different IA estimation methods.

6.1. Method

The same DGMs as in Simulation Study 1 were employed to simulate the data (see Section 5). The standard error computation described in Section 3 was applied for all estimators used in Simulation Study 1. Confidence intervals at a confidence level of 95% were computed using a normal distribution approximation (i.e., the estimated confidence interval was $\hat{\theta} \pm 1.96 \times SE(\hat{\theta})$). The coverage rate at a confidence level of 95% was computed as the percentage of the events that an estimated confidence interval covers for the true parameter value. Coverage rates were considered acceptable if they were not smaller than 91.0 or larger than 98.0 (see [56]).

In total, 5000 replications were conducted in each cell of the simulation. The IA method with standard error estimates of model parameters was again estimated with the `sirt::invariance.alignment()` function that is contained in the R [52] package `sirt` (Version 4.0-19; [55]). The R code used for this simulation can be found at <https://osf.io/7kwqh/> (accessed on 17 September 2023).

6.2. Results

Table 3 shows coverage rates for factor means μ_2 and μ_3 and factor SDs σ_2 and σ_3 . Overall, the coverage rates were acceptable. Undercoverage was observed when parameter estimates were biased (e.g., $p = 0.5, \epsilon = 0.01$). For approximately unbiased point estimates, standard errors based on the delta method can be reliably used. There is no need to rely on computationally more demanding standard error estimates like jackknife or bootstrap. Interestingly, the coverage rates were also satisfactory for the newly proposed loss function with the power $p = 0$.

Table 3. Simulation Study 2: coverage rates for factor means μ_2 and μ_3 and factor standard deviations σ_2 and σ_3 as a function of sample size N for different estimation methods.

Meth	p, ϵ	μ_2				μ_3				σ_2				σ_3			
		N				N				N				N			
		250	500	1000	2000	250	500	1000	2000	250	500	1000	2000	250	500	1000	2000
<i>Noninvariant Item Intercepts (DGM1)</i>																	
NOL	0.5, 0.01	92.7	91.4	88.0	84.4	92.5	90.6	89.2	85.7	95.6	95.5	95.5	95.2	95.8	95.4	95.4	94.9
	0.5, 0.001	95.8	95.6	94.8	94.9	95.9	95.9	95.8	95.0	97.5	97.4	96.8	96.1	98.1	97.3	96.9	96.0
	0.25, 0.01	94.5	93.6	91.6	90.0	94.4	93.1	92.5	90.9	95.8	95.6	95.6	95.2	96.1	95.7	95.5	95.0
	0.25, 0.001	96.2	96.4	95.8	95.7	96.5	96.8	96.5	96.0	97.8	97.4	97.1	96.5	98.2	97.6	97.3	96.3
	0.1, 0.01	94.9	94.2	92.7	92.2	95.2	94.0	93.5	92.5	96.1	95.7	95.6	95.3	96.4	95.8	95.6	95.0
	0.1, 0.001	96.2	96.5	96.0	96.0	96.6	97.1	96.7	96.2	97.9	97.5	97.3	96.6	98.0	97.6	97.5	96.5
	0, 0.01	96.3	96.2	95.3	95.6	96.5	96.6	95.9	95.5	97.2	96.9	96.4	95.6	97.4	96.7	96.5	95.5
	0, 0.001	96.2	96.7	96.5	96.6	96.8	97.4	97.3	97.3	97.6	97.5	97.6	97.6	97.9	97.7	97.4	97.1
LOG	0.5, 0.01	92.5	91.3	87.8	84.3	92.0	90.3	88.9	85.4	95.6	95.3	95.5	95.1	95.6	95.3	95.4	95.0
	0.5, 0.001	95.7	95.5	94.8	94.8	95.7	95.7	95.7	94.9	97.5	97.1	96.8	96.1	97.8	97.2	96.9	96.0
	0.25, 0.01	94.3	93.5	91.5	90.0	93.9	92.9	92.2	90.7	95.7	95.3	95.6	95.2	95.9	95.4	95.5	95.0
	0.25, 0.001	95.9	96.3	95.8	95.7	96.2	96.6	96.5	95.9	97.6	97.4	97.1	96.4	97.8	97.5	97.2	96.3
	0.1, 0.01	94.7	94.1	92.6	92.2	94.7	93.8	93.3	92.4	95.9	95.4	95.6	95.2	96.0	95.5	95.5	95.0
	0.1, 0.001	96.0	96.5	96.0	96.0	96.4	97.1	96.7	96.1	97.6	97.4	97.2	96.6	97.9	97.5	97.3	96.4
	0, 0.01	96.3	96.2	95.3	95.6	96.4	96.6	95.8	95.5	97.0	96.7	96.3	95.5	97.0	96.7	96.5	95.5
	0, 0.001	96.2	96.7	96.4	96.7	96.7	97.3	97.3	97.3	97.7	97.4	97.7	97.4	97.5	97.5	97.4	97.0

Table 3. Cont.

Meth	p, ϵ	μ_2				μ_3				σ_2				σ_3			
		N				N				N				N			
		250	500	1000	2000	250	500	1000	2000	250	500	1000	2000	250	500	1000	2000
<i>Noninvariant Item Intercepts and Noninvariant Item Loadings (DGM2)</i>																	
NOL	0.5, 0.01	96.0	95.6	95.1	94.8	94.5	94.1	93.2	91.7	84.9	78.6	64.4	41.7	96.2	94.6	93.4	91.6
	0.5, 0.001	96.9	96.7	96.0	96.1	96.6	96.6	96.2	96.1	92.2	95.4	95.2	93.0	97.3	96.9	96.6	96.2
	0.25, 0.01	96.4	95.9	95.4	95.2	95.6	95.1	94.3	93.8	89.0	89.1	81.8	69.5	96.6	95.4	94.6	93.7
	0.25, 0.001	96.7	96.7	96.4	96.3	96.7	96.9	96.9	96.5	91.0	95.4	96.0	95.3	97.2	96.8	96.9	96.4
	0.1, 0.01	96.4	96.0	95.3	95.3	95.9	95.4	94.9	94.4	89.7	92.4	86.9	79.1	96.6	95.6	94.9	94.2
	0.1, 0.001	96.8	96.5	96.4	96.4	96.7	96.9	97.0	96.7	90.6	95.5	96.5	96.0	96.9	96.8	96.8	96.6
	0, 0.01	95.9	96.4	95.6	95.6	95.7	96.8	96.2	95.6	87.5	94.7	96.0	95.3	95.3	96.3	96.1	95.8
	0, 0.001	95.9	96.4	96.6	96.7	95.7	96.9	97.2	97.4	86.5	94.6	96.9	96.6	95.5	96.3	97.0	96.9
LOG	0.5, 0.01	95.3	94.9	93.8	93.4	91.8	91.0	89.2	85.6	92.7	88.7	81.6	70.1	96.0	95.5	95.2	95.2
	0.5, 0.001	96.6	96.5	95.9	95.8	95.5	96.2	95.9	95.7	96.5	96.9	95.9	94.8	97.5	97.0	96.7	96.2
	0.25, 0.01	95.8	95.4	94.6	94.3	93.7	93.3	92.5	90.8	94.8	93.3	88.9	83.4	96.2	95.8	95.3	95.2
	0.25, 0.001	96.6	96.6	96.2	96.2	95.9	96.7	96.5	96.5	96.4	97.0	96.7	96.0	97.2	97.0	96.9	96.5
	0.1, 0.01	96.0	95.8	94.9	94.7	94.5	94.3	93.4	92.7	95.5	94.8	91.4	87.7	96.3	96.0	95.4	95.2
	0.1, 0.001	96.7	96.6	96.3	96.4	95.9	96.7	96.8	96.7	96.3	96.9	97.0	96.3	96.9	97.1	97.0	96.6
	0, 0.01	96.1	96.5	95.6	95.5	95.3	96.8	96.1	95.5	95.2	96.6	96.1	95.3	95.7	96.5	96.1	95.8
	0, 0.001	96.2	96.6	96.6	96.7	95.1	97.0	97.1	97.3	94.7	96.5	96.8	96.5	96.0	96.6	96.8	96.9

Note. Meth = estimation method; NOL = no logarithmized linking function for item loadings using H_1 in (8); LOG = logarithmized linking function for item loadings using H_1^* in (10); p = power used in the L_p loss function ρ ; ϵ = tuning parameter used in the differentiable approximation of the L_p loss function ρ ; coverage rates smaller than 91.0 or larger than 98.0 are shown with a gray background.

7. Simulation Study 3: Bias and RMSE in a Six-Group Example

In the last example, Simulation Study 3, different estimation methods of IA were examined for DGMs involving six groups and four items.

7.1. Method

The data were simulated from a one-dimensional factor model involving four items (i.e., $I = 4$) and six groups (i.e., $G = 6$). The factor variable was normally distributed with group means 0, -0.27, -0.46, 0.11, 0.21, and 0.49, and the group variances were 1, 0.95, 0.87, 1.23, 1.1, and 0.99, respectively. All measurement error variances were set to one in all groups and uncorrelated with each other. The factor variable and residual variables were normally distributed.

Two DGMs were simulated that refer to a violation of measurement invariance. Invariance only appeared in item loadings, while item intercepts had invariant parameters.

In the first DGM of Simulation Study 3 (i.e., DGM3), the DIF effects in item loadings λ_{ig} ($i = 1, \dots, I, \dots$) were unidirectional. That is, item loadings with DIF loadings were all larger than one, while the invariant item parameters had loadings equal to one. The item loadings with DIF effects were as follows: $\lambda_{12} = 2.014$ (i.e., first item in second group), $\lambda_{21} = 1.733$, $\lambda_{23} = 2.014$, and $\lambda_{36} = 2.117$. The item intercepts were all set to 0 in DGM3.

In the second DGM of Simulation Study 3 (i.e., DGM4), the DIF effects in item loadings were directional. That is, item loadings with DIF effects could be smaller or larger than one. The item loadings with DIF effects were as follows: $\lambda_{12} = 0.497$, $\lambda_{21} = 1.733$, $\lambda_{23} = 2.014$, and $\lambda_{36} = 2.117$. Like in DGM3, the item intercepts were all set to 0 in DGM4.

As in Simulation Study 1 and Simulation Study 2, the sample size per group was chosen as $N = 250$, $N = 500$, $N = 1000$, or $N = 2000$. The same IA estimation methods as in the other two studies were utilized. To summarize the performance of the parameter estimates across the six groups, average absolute bias, average SD, and the average relative RMSE were computed, where the average was calculated for factor means μ and factor SDs σ , separately.

Again, the simulation study was carried out in the R [52] software. IA was estimated using the `sirt::invariance.alignment()` function in the R package `sirt` (Version 4.0-19; [55]). Replication material can be found at <https://osf.io/7kwqh/> (accessed on 17 September 2023).

7.2. Results

Table 4 displays average absolute bias, average SD, and average relative RMSE for factor means μ and factor SDs σ for the DGM with unidirectional effects of noninvariance (i.e., DGM3). As expected from Simulation Study 1, noninvariant item loadings mainly impacted factor standard deviations. It was obtained that factor SDs had a larger absolute bias for the (default) NOL method compared with the LOG method. However, absolute bias decreased with larger sample sizes and by choosing the tuning parameter $\epsilon = 0.001$ instead of $\epsilon = 0.01$. Furthermore, the LOG method had a much smaller bias in estimated factor SDs. This finding also translated into findings for the average absolute RMSE of the σ estimates. Interestingly, the LOG method was also preferred for the estimated factor means. The LOG estimates for μ had, on average, smaller SDs than the NOL estimates.

Similar findings were obtained in the case of the bidirectional DIF effects in item loadings (DGM4) that are displayed in Table A2 in Appendix B. The LOG method was preferred over the NOL method regarding the estimation of factor SDs σ . Furthermore, LOG resulted in slightly better estimates than the NOL method for factor means μ . IA with $p = 0.5$ should preferably choose the tuning parameter $\epsilon = 0.001$ instead of $\epsilon = 0.01$.

Table 4. Simulation Study 3: average absolute bias, average standard deviation (average SD), and average relative root mean square error (ARRMSE) for factor means μ and factor standard deviations σ as a function of sample size N for different estimation methods for unidirectional effects in noninvariance in item loadings (DGM3).

Meth	p, ϵ	Average Absolute Bias				Average SD				ARRMSE			
		N				N				N			
		250	500	1000	2000	250	500	1000	2000	250	500	1000	2000
<i>Factor Means μ</i>													
NOL	0.5, 0.01 [†]	0.047	0.030	0.025	0.021	0.108	0.071	0.050	0.034	100 [†]	100 [†]	100 [†]	100 [†]
	0.5, 0.001	0.034	0.017	0.011	0.007	0.108	0.071	0.048	0.034	95.1	92.9	87.4	83.2
	0.25, 0.01	0.033	0.019	0.015	0.011	0.104	0.070	0.048	0.034	91.4	92.1	89.9	87.7
	0.25, 0.001	0.025	0.011	0.007	0.003	0.107	0.071	0.048	0.033	91.9	91.4	84.9	81.0
	0.1, 0.01	0.028	0.015	0.012	0.008	0.102	0.069	0.048	0.033	88.9	89.9	87.1	84.6
	0.1, 0.001	0.022	0.009	0.005	0.002	0.107	0.071	0.048	0.033	91.6	90.9	85.1	81.1
	0, 0.01	0.012	0.004	0.003	0.001	0.105	0.069	0.047	0.033	88.4	87.1	82.2	79.1
	0, 0.001	0.012	0.005	0.003	0.001	0.111	0.074	0.051	0.035	93.5	93.7	90.1	84.7
LOG	0.5, 0.01	0.003	0.003	0.003	0.003	0.091	0.065	0.046	0.032	75.9	82.0	80.6	78.7
	0.5, 0.001	0.003	0.003	0.002	0.002	0.094	0.067	0.046	0.033	78.8	84.4	80.9	79.1
	0.25, 0.01	0.003	0.003	0.002	0.003	0.092	0.065	0.046	0.032	76.5	82.1	80.8	78.7
	0.25, 0.001	0.002	0.003	0.002	0.002	0.096	0.068	0.047	0.033	80.6	85.8	81.6	79.3
	0.1, 0.01	0.003	0.003	0.002	0.002	0.092	0.066	0.046	0.032	77.0	82.5	80.7	78.8
	0.1, 0.001	0.002	0.003	0.002	0.001	0.098	0.069	0.047	0.033	81.6	87.3	82.4	79.9
	0, 0.01	0.002	0.002	0.002	0.001	0.098	0.068	0.046	0.033	81.8	85.5	81.4	78.9
	0, 0.001	0.002	0.003	0.002	0.001	0.104	0.074	0.051	0.035	86.8	92.4	89.7	84.7

Table 4. Cont.

Meth	p, ϵ	Average Absolute Bias				Average SD				ARRMSE			
		N				N				N			
		250	500	1000	2000	250	500	1000	2000	250	500	1000	2000
<i>Factor SDs σ</i>													
NOL	0.5, 0.01 [†]	0.126	0.091	0.072	0.063	0.118	0.071	0.048	0.035	100 [†]	100 [†]	100 [†]	100 [†]
	0.5, 0.001	0.087	0.053	0.031	0.021	0.120	0.073	0.049	0.034	86.6	78.4	67.9	55.6
	0.25, 0.01	0.084	0.060	0.043	0.035	0.111	0.069	0.048	0.034	80.9	79.3	74.6	68.8
	0.25, 0.001	0.064	0.037	0.019	0.011	0.116	0.073	0.049	0.033	77.4	71.2	61.0	49.4
	0.1, 0.01	0.071	0.049	0.033	0.026	0.109	0.068	0.047	0.034	75.6	73.0	67.1	60.0
	0.1, 0.001	0.056	0.031	0.015	0.008	0.117	0.073	0.049	0.033	75.4	68.8	59.3	48.1
	0, 0.01	0.031	0.018	0.008	0.004	0.110	0.069	0.046	0.032	66.7	61.9	54.6	45.4
	0, 0.001	0.028	0.016	0.006	0.002	0.118	0.077	0.052	0.035	70.5	68.3	60.8	49.8
LOG	0.5, 0.01	0.015	0.010	0.009	0.007	0.090	0.062	0.044	0.032	53.3	54.7	52.1	46.4
	0.5, 0.001	0.010	0.006	0.004	0.002	0.097	0.066	0.045	0.032	56.8	57.7	53.0	45.7
	0.25, 0.01	0.011	0.007	0.006	0.005	0.091	0.062	0.044	0.032	53.4	54.6	51.8	45.6
	0.25, 0.001	0.008	0.004	0.002	0.001	0.100	0.068	0.046	0.032	58.1	59.1	53.6	46.0
	0.1, 0.01	0.009	0.006	0.005	0.004	0.092	0.063	0.044	0.032	53.8	54.6	51.7	45.3
	0.1, 0.001	0.006	0.005	0.002	0.001	0.101	0.069	0.047	0.033	59.0	60.1	54.3	46.2
	0, 0.01	0.003	0.007	0.004	0.002	0.101	0.067	0.046	0.032	58.8	58.6	53.2	45.4
	0, 0.001	0.003	0.006	0.002	0.001	0.108	0.076	0.052	0.035	63.0	66.4	60.3	49.9

Note. Meth = estimation method; NOL = no logarithmized linking function for item loadings using H_1 in (8); LOG = logarithmized linking function for item loadings using H_1^* in (10); p = power used in the L_p loss function; ρ ; ϵ = tuning parameter used in the differentiable approximation of the L_p loss function; values of average absolute bias larger than 0.03 are shown with a gray background. [†]The Mplus defaults $p = 0.5$, $\epsilon = 0.01$, and NOL are the reference methods in the computation of the relative RMSE. ARRMSE values smaller than 95.0 are printed in bold font.

8. Discussion and Conclusions

In this article, we critically discussed implementation aspects in IA. Because IA is now widely applied in the social sciences, researchers should opt for appropriate estimation methods. We derived recommendations for software implementation and practical application of IA through three simulation studies.

In IA, the loss function $\rho(x) = \sqrt{|x|} = |x|^{0.5}$ is the default choice in the popular Mplus software. A differentiable approximation of this loss function uses the tuning parameter $\epsilon = 0.01$ as a default in this software. Our simulations revealed that this default choice can induce bias in estimated factor means and factor standard deviations. The bias can be reduced by switching to the tuning parameter $\epsilon = 0.001$. Notably, biases in IA were particularly pronounced in small to moderate samples (i.e., $N = 250$ persons per group). It turned out that bias in estimated factor standard deviations occurred in the presence of noninvariant item loadings. This bias can be reduced by using a modified IA optimization function in which logarithmized item loadings are aligned (i.e., the LOG method described in this paper). In general, we found that the LOG method generally improves the default NOL method in IA (which uses no logarithmized item loadings) in the situations in which bias occurred and performed comparably to NOL in all other situations. Furthermore, the L_0 loss function recently proposed by O'Neill and Burke (i.e., $p = 0$) showed the least bias across all simulated conditions. This method can be regarded as the frontrunner across all simulation conditions when used with the tuning parameter $\epsilon = 0.01$. Finally, statistical inference based on the delta method performed satisfactorily for all approximately unbiased estimates in terms of coverage rates.

In future research, the generalizability of the findings of this study to more groups or more items [17] can be examined. Moreover, implementation aspects of IA could also be investigated for dichotomous or ordinal items [19,20]. In particular, the performance of IA in small samples [57] requires additional consideration. It might be that regularized estimation [58–61] or confirmatory factor analysis estimation that uses robust loss functions (i.e., model-robust estimation; see [62,63]) have advantages over IA in small samples.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

- CFA confirmatory factor analysis;
- DGM data-generating model;
- DIF differential item functioning;
- IA invariance alignment;
- ML maximum likelihood;
- PISA programme for international student assessment;
- RMSE root mean square error;
- SD standard deviation;
- TIMSS trends in international mathematics and science study.

Appendix A. Additional Results for Simulation Study 1

Table A1 displays bias, SD, and relative RMSE for the factor mean μ_3 and the factor SD σ_2 of the third group in the condition of noninvariant item loadings and noninvariant item intercepts (DGM2).

Table A1. Simulation Study 1: bias, standard deviation (SD), and relative root mean square error (RMSE) of the factor mean μ_3 and the factor standard deviation σ_3 as a function of sample size N for different estimation methods in the case of noninvariant item intercepts and noninvariant item loadings (DGM2).

Meth	p, ϵ	Bias				SD				RMSE			
		N				N				N			
		250	500	1000	2000	250	500	1000	2000	250	500	1000	2000
<i>Factor Mean μ_3</i>													
NOL	0.5, 0.01 [†]	-0.037	-0.031	-0.027	-0.024	0.131	0.090	0.061	0.042	100 [†]	100 [†]	100 [†]	100 [†]
	0.5, 0.001	-0.019	-0.016	-0.011	-0.009	0.142	0.098	0.062	0.043	105.4	103.6	95.4	90.0
	0.25, 0.01	-0.022	-0.019	-0.018	-0.016	0.136	0.091	0.061	0.042	101.3	97.7	95.1	92.3
	0.25, 0.001	-0.006	-0.008	-0.006	-0.005	0.148	0.101	0.064	0.043	108.8	105.7	97.0	89.5
	0.1, 0.01	-0.015	-0.014	-0.013	-0.012	0.139	0.092	0.061	0.042	102.9	97.2	93.7	89.8
	0.1, 0.001	-0.002	-0.005	-0.004	-0.003	0.151	0.101	0.065	0.043	111.2	105.3	97.5	89.7
	0, 0.01	0.016	0.007	0.000	0.000	0.163	0.098	0.062	0.041	120.1	103.2	93.9	85.4
	0, 0.001	0.016	0.008	0.001	0.000	0.169	0.107	0.070	0.048	125.0	111.9	104.8	98.2
LOG	0.5, 0.01	-0.066	-0.053	-0.044	-0.038	0.126	0.087	0.059	0.041	104.4	106.3	110.6	115.8
	0.5, 0.001	-0.045	-0.030	-0.020	-0.014	0.136	0.094	0.061	0.043	105.5	103.4	97.0	92.1
	0.25, 0.01	-0.049	-0.038	-0.030	-0.026	0.130	0.088	0.059	0.041	102.6	100.0	100.4	100.5
	0.25, 0.001	-0.032	-0.020	-0.012	-0.008	0.141	0.096	0.063	0.043	106.6	102.8	96.9	90.0
	0.1, 0.01	-0.041	-0.031	-0.025	-0.021	0.133	0.088	0.060	0.041	102.7	97.9	97.0	95.2
	0.1, 0.001	-0.026	-0.015	-0.009	-0.006	0.145	0.097	0.064	0.043	108.5	102.9	97.0	89.7
	0, 0.01	-0.006	-0.001	-0.002	-0.002	0.154	0.095	0.062	0.041	113.1	98.9	93.1	85.2
	0, 0.001	-0.006	0.000	-0.002	-0.001	0.160	0.103	0.069	0.047	117.4	107.6	104.0	97.7

Table A1. Cont.

Meth	p, ϵ	Bias				SD				RMSE			
		N				N				N			
		250	500	1000	2000	250	500	1000	2000	250	500	1000	2000
<i>Factor SD σ_3</i>													
NOL	0.5, 0.01 [†]	0.051	0.034	0.028	0.022	0.102	0.069	0.048	0.033	100 [†]	100 [†]	100 [†]	100 [†]
	0.5, 0.001	0.046	0.023	0.015	0.009	0.116	0.075	0.049	0.033	109.5	100.7	92.7	85.5
	0.25, 0.01	0.048	0.029	0.022	0.016	0.107	0.071	0.048	0.033	103.0	98.5	95.3	92.3
	0.25, 0.001	0.045	0.020	0.011	0.006	0.122	0.078	0.050	0.033	114.1	103.2	92.3	83.8
	0.1, 0.01	0.047	0.026	0.019	0.014	0.109	0.071	0.048	0.033	104.4	98.0	93.3	89.3
	0.1, 0.001	0.044	0.018	0.010	0.005	0.124	0.079	0.050	0.033	115.1	104.2	92.8	83.5
	0, 0.01	0.040	0.014	0.006	0.003	0.132	0.079	0.048	0.032	121.4	103.2	87.3	80.1
	0, 0.001	0.040	0.015	0.007	0.003	0.138	0.086	0.054	0.035	126.4	112.7	98.4	89.2
LOG	0.5, 0.01	0.007	0.002	0.003	0.001	0.097	0.066	0.046	0.032	85.5	85.4	83.8	80.7
	0.5, 0.001	0.008	0.003	0.003	0.002	0.109	0.071	0.048	0.032	95.7	92.1	86.9	81.6
	0.25, 0.01	0.007	0.003	0.003	0.001	0.102	0.067	0.047	0.032	89.8	86.9	84.6	80.8
	0.25, 0.001	0.008	0.003	0.003	0.002	0.114	0.073	0.049	0.033	100.4	94.8	88.1	81.8
	0.1, 0.01	0.008	0.003	0.003	0.001	0.105	0.068	0.047	0.032	92.0	87.7	84.8	80.8
	0.1, 0.001	0.008	0.004	0.003	0.002	0.117	0.074	0.049	0.033	102.8	96.1	88.9	82.0
	0, 0.01	0.010	0.003	0.003	0.001	0.124	0.074	0.047	0.032	108.9	94.9	85.6	79.5
	0, 0.001	0.011	0.004	0.003	0.002	0.128	0.081	0.053	0.036	112.9	104.1	96.6	89.3

Note. Meth = estimation method; NOL = no logarithmized linking function for item loadings using H_1 in (8); LOG = logarithmized linking function for item loadings using H_1^* in (10); p = power used in the L_p loss function; ϵ = tuning parameter used in the differentiable approximation of the L_p loss function; absolute biases larger than 0.03 are shown with a gray background. [†] The Mplus defaults $p = 0.5, \epsilon = 0.01$, and NOL are the reference methods in the computation of the relative RMSE. Relative RMSE values smaller than 95.0 are printed in bold font.

Appendix B. Additional Results for Simulation Study 3

Table A2 displays the average absolute bias, average SD, and average relative RMSE for factor means and factor SDs for the DGM4, which contained bidirectional DIF effects in item loadings.

Table A2. Simulation Study 3: average absolute bias, average standard deviation (average SD), and average relative root mean square error (ARRMSE) for factor means μ and factor standard deviations σ as a function of sample size N for different estimation methods for bidirectional effects in noninvariance in item loadings (DGM4).

Meth	p, ϵ	Average Absolute Bias				Average SD				ARRMSE			
		N				N				N			
		250	500	1000	2000	250	500	1000	2000	250	500	1000	2000
<i>Factor Means μ</i>													
NOL	0.5, 0.01 [†]	0.029	0.023	0.019	0.016	0.105	0.073	0.051	0.036	100 [†]	100 [†]	100 [†]	100 [†]
	0.5, 0.001	0.020	0.014	0.009	0.006	0.107	0.073	0.051	0.035	99.1	95.1	93.0	88.0
	0.25, 0.01	0.021	0.016	0.012	0.010	0.103	0.072	0.051	0.035	95.7	95.0	94.7	91.8
	0.25, 0.001	0.015	0.010	0.006	0.003	0.109	0.074	0.051	0.035	99.2	94.5	91.7	86.3
	0.1, 0.01	0.018	0.013	0.010	0.008	0.103	0.072	0.050	0.035	94.6	93.3	92.9	89.4
	0.1, 0.001	0.013	0.009	0.004	0.003	0.109	0.074	0.051	0.035	99.2	94.2	91.7	86.6
	0, 0.01	0.009	0.005	0.002	0.002	0.110	0.073	0.050	0.035	98.7	92.0	89.2	85.1
	0, 0.001	0.010	0.004	0.002	0.001	0.116	0.077	0.054	0.037	104.5	97.9	96.6	91.7

Table A2. Cont.

Meth	p, ϵ	Average Absolute Bias				Average SD				ARRMSE			
		N				N				N			
		250	500	1000	2000	250	500	1000	2000	250	500	1000	2000
LOG	0.5, 0.01	0.008	0.004	0.005	0.004	0.094	0.068	0.048	0.034	84.7	86.9	87.2	84.1
	0.5, 0.001	0.006	0.002	0.002	0.002	0.099	0.070	0.049	0.034	89.2	88.2	87.9	84.2
	0.25, 0.01	0.006	0.003	0.003	0.002	0.094	0.069	0.048	0.034	84.9	87.3	87.3	84.0
	0.25, 0.001	0.005	0.002	0.001	0.001	0.102	0.071	0.049	0.035	91.7	89.7	88.6	85.0
	0.1, 0.01	0.005	0.003	0.003	0.002	0.096	0.069	0.049	0.034	85.9	87.3	87.4	84.0
	0.1, 0.001	0.005	0.002	0.001	0.001	0.103	0.072	0.050	0.035	92.7	90.5	89.0	85.6
	0, 0.01	0.003	0.002	0.001	0.001	0.105	0.071	0.049	0.035	93.8	89.9	88.4	84.7
	0, 0.001	0.004	0.001	0.001	0.001	0.110	0.076	0.053	0.037	98.5	95.6	95.7	91.8
		Factor SDs σ											
NOL	0.5, 0.01 [†]	0.083	0.062	0.050	0.042	0.099	0.067	0.048	0.033	100 [†]	100 [†]	100 [†]	100 [†]
	0.5, 0.001	0.061	0.039	0.023	0.014	0.104	0.071	0.050	0.033	93.7	89.6	80.3	69.3
	0.25, 0.01	0.062	0.044	0.033	0.026	0.098	0.067	0.048	0.033	89.8	88.0	84.7	79.6
	0.25, 0.001	0.047	0.029	0.014	0.008	0.105	0.073	0.049	0.033	89.6	86.3	76.2	65.6
	0.1, 0.01	0.054	0.037	0.026	0.020	0.098	0.067	0.047	0.033	87.3	84.3	79.6	73.5
	0.1, 0.001	0.043	0.025	0.011	0.006	0.106	0.073	0.050	0.033	89.0	85.3	75.5	64.9
	0, 0.01	0.030	0.015	0.007	0.003	0.105	0.069	0.047	0.032	85.2	78.7	70.0	62.3
	0, 0.001	0.029	0.013	0.005	0.002	0.113	0.078	0.053	0.036	90.9	87.4	79.2	70.9
LOG	0.5, 0.01	0.029	0.023	0.018	0.016	0.089	0.063	0.044	0.031	76.7	77.2	75.0	74.3
	0.5, 0.001	0.019	0.012	0.008	0.006	0.097	0.067	0.046	0.032	78.7	76.4	70.4	64.4
	0.25, 0.01	0.021	0.015	0.012	0.010	0.091	0.063	0.045	0.031	74.8	73.5	70.2	66.4
	0.25, 0.001	0.013	0.007	0.005	0.003	0.100	0.069	0.047	0.032	80.1	76.9	70.7	63.8
	0.1, 0.01	0.017	0.012	0.009	0.007	0.091	0.064	0.045	0.031	74.2	72.6	68.9	63.9
	0.1, 0.001	0.011	0.005	0.003	0.003	0.102	0.069	0.048	0.032	80.9	77.5	71.1	63.9
	0, 0.01	0.004	0.002	0.001	0.001	0.101	0.068	0.046	0.031	79.4	74.9	68.4	61.7
	0, 0.001	0.004	0.001	0.001	0.001	0.109	0.076	0.053	0.036	85.8	84.2	78.4	70.6

Note. Meth = estimation method; NOL = no logarithmized linking function for item loadings using H_1 in (8); LOG = logarithmized linking function for item loadings using H_1^* in (10); p = power used in the L_p loss function ρ ; ϵ = tuning parameter used in the differentiable approximation of the L_p loss function ρ ; values of average absolute bias larger than 0.03 are shown with a gray background. [†] The Mplus defaults $p = 0.5$, $\epsilon = 0.01$, and NOL are the reference methods in the computation of the relative RMSE. ARRMSE values smaller than 95.0 are printed in bold font.

References

1. Meredith, W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika* **1993**, *58*, 525–543. [CrossRef]
2. Mellenbergh, G.J. Item bias and item response theory. *Int. J. Educ. Res.* **1989**, *13*, 127–143. [CrossRef]
3. Millsap, R.E. *Statistical Approaches to Measurement Invariance*; Routledge: New York, NY, USA, 2011. [CrossRef]
4. van de Vijver, F.J.R. (Ed.) *Invariance Analyses in Large-Scale Studies*; OECD: Paris, France, 2019. [CrossRef]
5. Asparouhov, T.; Muthén, B. Multiple-group factor analysis alignment. *Struct. Equ. Model.* **2014**, *21*, 495–508. [CrossRef]
6. Muthén, B.; Asparouhov, T. IRT studies of many groups: The alignment method. *Front. Psychol.* **2014**, *5*, 978. [CrossRef]
7. Arts, I.; Fang, Q.; Meitinger, K.; van de Schoot, R. Approximate measurement invariance of willingness to sacrifice for the environment across 30 countries: The importance of prior distributions and their visualization. *Front. Psychol.* **2021**, *12*, 624032. [CrossRef]
8. Ciecuch, J.; Davidov, E.; Schmidt, P. Alignment optimization. Estimation of the most trustworthy means in cross-cultural studies even in the presence of noninvariance. In *Cross-Cultural Analysis: Methods and Applications*; Davidov, E., Schmidt, P., Billiet, J., Eds.; Routledge: London, UK, 2018; pp. 571–592. [CrossRef]
9. Pokropek, A.; Davidov, E.; Schmidt, P. A Monte Carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Struct. Equ. Model.* **2019**, *26*, 724–744. [CrossRef]

10. Leitgöb, H.; Seddig, D.; Asparouhov, T.; Behr, D.; Davidov, E.; De Roover, K.; Jak, S.; Meitinger, K.; Menold, N.; Muthén, B.; et al. Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Soc. Sci. Res.* **2023**, *110*, 102805. [[CrossRef](#)]
11. Luong, R.; Flake, J.K. Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychol. Methods* **2023**, *28*, 905–924. [[CrossRef](#)]
12. Pokropek, A.; Lüdtke, O.; Robitzsch, A. An extension of the invariance alignment method for scale linking. *Psych. Test Assess. Model.* **2020**, *62*, 303–334. Available online: <https://bit.ly/2UEp9GH> (accessed on 17 September 2023).
13. Mansolf, M.; Vreeker, A.; Reise, S.P.; Freimer, N.B.; Glahn, D.C.; Gur, R.E.; Moore, T.M.; Pato, C.N.; Pato, M.T.; Palotie, A.; et al. Extensions of multiple-group item response theory alignment: Application to psychiatric phenotypes in an international genomics consortium. *Educ. Psychol. Meas.* **2020**, *80*, 870–909. [[CrossRef](#)]
14. Muthén, L.; Muthén, B. *Mplus User's Guide*; Muthén & Muthén: Los Angeles, CA, USA, 1998–2023.
15. Kim, E.S.; Cao, C.; Wang, Y.; Nguyen, D.T. Measurement invariance testing with many groups: A comparison of five approaches. *Struct. Equ. Model.* **2017**, *24*, 524–544. [[CrossRef](#)]
16. Lai, M.H.C.; Liu, Y.; Tse, W.W.Y. Adjusting for partial invariance in latent parameter estimation: Comparing forward specification search and approximate invariance methods. *Behav. Res. Methods* **2022**, *54*, 414–434. [[CrossRef](#)] [[PubMed](#)]
17. Muthén, B.; Asparouhov, T. Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociol. Methods Res.* **2018**, *47*, 637–664. [[CrossRef](#)]
18. DeMars, C.E. Alignment as an alternative to anchor purification in DIF analyses. *Struct. Equ. Model.* **2020**, *27*, 56–72. [[CrossRef](#)]
19. Finch, W.H. Detection of differential item functioning for more than two groups: A Monte Carlo comparison of methods. *Appl. Meas. Educ.* **2016**, *29*, 30–45. [[CrossRef](#)]
20. Flake, J.K.; McCoach, D.B. An investigation of the alignment method with polytomous indicators under conditions of partial measurement invariance. *Struct. Equ. Model.* **2018**, *25*, 56–70. [[CrossRef](#)]
21. Byrne, B.M.; van de Vijver, F.J.R. The maximum likelihood alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema* **2017**, *29*, 539–551. [[CrossRef](#)]
22. Marsh, H.W.; Guo, J.; Parker, P.D.; Nagengast, B.; Asparouhov, T.; Muthén, B.; Dicke, T. What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychol. Methods* **2018**, *23*, 524–545. [[CrossRef](#)] [[PubMed](#)]
23. Kim, E.; Cao, C.; Liu, S.; Wang, Y.; Dedrick, R. Testing measurement invariance over time with intensive longitudinal data and identifying a source of non-invariance. *Struct. Equ. Model.* **2023**, *30*, 393–411. [[CrossRef](#)]
24. Lai, M.H.C. Adjusting for measurement noninvariance with alignment in growth modeling. *Multivar. Behav. Res.* **2023**, *58*, 30–47. [[CrossRef](#)]
25. Seddig, D.; Leitgöb, H. Approximate measurement invariance and longitudinal confirmatory factor analysis: Concept and application with panel data. *Surv. Res. Methods* **2018**, *12*, 29–41. [[CrossRef](#)]
26. Winter, S.D.; Depaoli, S. An illustration of Bayesian approximate measurement invariance with longitudinal data and a small sample size. *Int. J. Behav. Dev.* **2020**, *44*, 371–382. [[CrossRef](#)]
27. Asparouhov, T.; Muthén, B. Multiple group alignment for exploratory and structural equation models. *Struct. Equ. Model.* **2023**, *30*, 169–191. [[CrossRef](#)]
28. Asparouhov, T.; Muthén, B. *Penalized Structural Equation Models*; Technical Report. 2023. Available online: <https://rb.gy/tbaj7> (accessed on 28 March 2023).
29. Davidov, E.; Cieciuch, J.; Meuleman, B.; Schmidt, P.; Algesheimer, R.; Hausherr, M. The comparability of measurements of attitudes toward immigration in the European Social Survey: Exact versus approximate measurement equivalence. *Public Opin. Q.* **2015**, *79*, 244–266. [[CrossRef](#)]
30. Munck, I.; Barber, C.; Torney-Purta, J. Measurement invariance in comparing attitudes toward immigrants among youth across Europe in 1999 and 2009: The alignment method applied to IEA CIVED and ICCS. *Sociol. Methods Res.* **2018**, *47*, 687–728. [[CrossRef](#)]
31. Caycho-Rodríguez, T.; Vilca, L.W.; Cervigni, M.; Gallegos, M.; Martino, P.; Calandra, M.; Rey Anacona, C.A.; López-Calle, C.; Moreta-Herrera, R.; Chacón-Andrade, E.R.; et al. Cross-national measurement invariance of the purpose in life test in seven Latin American countries. *Front. Psychol.* **2022**, *13*, 974133. [[CrossRef](#)]
32. Sideridis, G.; Alahmadi, M. Bullying in elementary schools: Differences across countries in the Persian gulf. *Children* **2023**, *10*, 1108. [[CrossRef](#)]
33. Sideridis, G.; Alghamdi, M.H. Bullying in middle school: Evidence for a multidimensional structure and measurement invariance across gender. *Children* **2023**, *10*, 873. [[CrossRef](#)]

34. Ding, Y.; Yang Hansen, K.; Klapp, A. Testing measurement invariance of mathematics self-concept and self-efficacy in PISA using MGCFAs and the alignment method. *Eur. J. Psychol. Educ.* **2023**, *38*, 709–732. [CrossRef]
35. Sirganci, G.; Uyumaz, G.; Yandi, A. Measurement invariance testing with alignment method: Many groups comparison. *Int. J. Assess. Tool. Educ.* **2020**, *7*, 657–673. [CrossRef]
36. Wurster, S. Measurement invariance of non-cognitive measures in TIMSS across countries and across time. An application and comparison of multigroup confirmatory factor analysis, Bayesian approximate measurement invariance and alignment optimization approach. *Stud. Educ. Eval.* **2022**, *73*, 101143. [CrossRef]
37. De Bondt, N.; Van Petegem, P. Psychometric evaluation of the overexcitability questionnaire-two applying Bayesian structural equation modeling (BSEM) and multiple-group BSEM-based alignment with approximate measurement invariance. *Front. Psychol.* **2015**, *6*, 1963. [CrossRef] [PubMed]
38. Wickham, R.E.; Gutierrez, R.; Giordano, B.L.; Rostosky, S.S.; Riggle, E.D.B. Gender and generational differences in the internalized homophobia questionnaire: An alignment IRT analysis. *Assessment* **2021**, *28*, 1159–1172. [CrossRef] [PubMed]
39. Eryilmaz, N.; Sandoval-Hernandez, A. Is distributed leadership universal? A cross-cultural, comparative approach across 40 Countries: An alignment optimisation approach. *Educ. Sci.* **2023**, *13*, 218. [CrossRef]
40. Lomazzi, V. Using alignment optimization to test the measurement invariance of gender role attitudes in 59 countries. *Methods Data Anal.* **2018**, *12*, 77–103. [CrossRef]
41. Bartholomew, D.J.; Knott, M.; Moustaki, I. *Latent Variable Models and Factor Analysis: A Unified Approach*; Wiley: New York, NY, USA, 2011. [CrossRef]
42. Holland, P.W.; Wainer, H. (Eds.) *Differential Item Functioning: Theory and Practice*; Lawrence Erlbaum: Hillsdale, NJ, USA, 1993. [CrossRef]
43. van de Schoot, R.; Kluytmans, A.; Tummers, L.; Lugtig, P.; Hox, J.; Muthén, B. Facing off with scylla and charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Front. Psychol.* **2013**, *4*, 770. [CrossRef]
44. Byrne, B.M.; Shavelson, R.J.; Muthén, B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychol. Bull.* **1989**, *105*, 456–466. [CrossRef]
45. Robitzsch, A. L_p loss functions in invariance alignment and Haberman linking with few or many groups. *Stats* **2020**, *3*, 246–283. [CrossRef]
46. Haberman, S.J. *Linking Parameter Estimates Derived from an Item Response Model through Separate Calibrations*; Research Report No. RR-09-40; Educational Testing Service: Princeton, NJ, USA, 2009. [CrossRef]
47. Davies, P.L. *Data Analysis and Approximate Models*; CRC Press: Boca Raton, FL, USA, 2014. [CrossRef]
48. Davies, P.L.; Terbeck, W. Interactions and outliers in the two-way analysis of variance. *Ann. Stat.* **1998**, *26*, 1279–1305. [CrossRef]
49. O’Neill, M.; Burke, K. Variable selection using a smooth information criterion for distributional regression models. *Stat. Comput.* **2023**, *33*, 71. [CrossRef]
50. Battauz, M. Regularized estimation of the nominal response model. *Multivar. Behav. Res.* **2020**, *55*, 811–824. [CrossRef] [PubMed]
51. Boos, D.D.; Stefanski, L.A. *Essential Statistical Inference*; Springer: New York, NY, USA, 2013. [CrossRef]
52. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2023. Available online: <https://www.R-project.org/> (accessed on 15 March 2023).
53. Fischer, R.; Karl, J.A. A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Front. Psychol.* **2019**, *10*, 1507. [CrossRef] [PubMed]
54. Han, H. Using measurement alignment in research on adolescence involving multiple groups: A brief tutorial with R. *J. Res. Adolesc.* **2023**. [CrossRef] [PubMed]
55. Robitzsch, A. *sirt: Supplementary Item Response Theory Models*; R Package Version 4.0-19. 2023. Available online: <https://github.com/alexanderrobitzsch/sirt> (accessed on 16 September 2023).
56. Muthén, L.K.; Muthén, B.O. How to use a Monte Carlo study to decide on sample size and determine power. *Struct. Equ. Model.* **2002**, *9*, 599–620. [CrossRef]
57. Wen, C.; Hu, F. Investigating the applicability of alignment—A Monte Carlo simulation study. *Front. Psychol.* **2022**, *13*, 845721. [CrossRef]
58. Huang, P.H. A penalized likelihood method for multi-group structural equation modelling. *Brit. J. Math. Stat. Psychol.* **2018**, *71*, 499–522. [CrossRef]
59. Jacobucci, R.; Grimm, K.J.; McArdle, J.J. Regularized structural equation modeling. *Struct. Equ. Model.* **2016**, *23*, 555–566. [CrossRef]
60. Qiao, X. Variable selection using L_q penalties. *WIREs Comput. Stat.* **2014**, *6*, 177–184. [CrossRef]
61. Robitzsch, A. Implementation aspects in regularized structural equation models. *Algorithms* **2023**, *16*, 446. [CrossRef]

-
62. Robitzsch, A. Comparing the robustness of the structural after measurement (SAM) approach to structural equation modeling (SEM) against local model misspecifications with alternative estimation approaches. *Stats* **2022**, *5*, 631–672. [[CrossRef](#)]
 63. Robitzsch, A. Model-robust estimation of multiple-group structural equation models. *Algorithms* **2023**, *16*, 210. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.