applied sciences

MDPI

Article

Implementation Aspects in Simulation Extrapolation-Based Stocking–Lord Linking

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: Stocking–Lord (SL) linking is commonly used for group comparisons based on dichotomous item responses. Recently, the simulation extrapolation (SIMEX) method for SL linking has been introduced within the two-parameter logistic (2PL) model to reduce the bias due to the presence of random differential item functioning (DIF). However, the original SIMEX approach can be computationally complex. This article evaluates two computational shortcuts to SIMEX-based SL linking through a simulation study: a quasi-Monte Carlo method that replicates items and a proposal using an approximate noniterative estimation instead of the iterative one in the SIMEX application of SL linking. The results demonstrate that these shortcuts yield linking parameter estimates with comparable statistical properties.

Keywords: Stocking–Lord linking; SIMEX; item response model; linking; 2PL model; approximation

check for updates

Academic Editor: Luigi Portinale

Received: 3 October 2024
Revised: 9 January 2025
Accepted: 15 January 2025
Published: 17 January 2025

Citation: Robitzsch, A. Implementation Aspects in Simulation Extrapolation-Based Stocking–Lord Linking. *Appl. Sci.* **2025**, *15*, 901. https://doi.org/10.3390/app15020901

Copyright: © 2025 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/).

1. Introduction

Item response theory (IRT) models [1,2] are statistical models to analyze multivariate dichotomous random variables. Let $\boldsymbol{X} = (X_1, \ldots, X_I)$ represent the vector of I dichotomous (i.e., binary) random variables $X_i \in \{0, 1\}$ (also referred to as items or item responses). A unidimensional IRT model [3] provides a statistical model for the probability distribution $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x})$ for $\boldsymbol{x} = (x_1, \ldots, x_I) \in \{0, 1\}^I$:

$$\mathsf{P}(\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\gamma}) = \int \prod_{i=1}^{I} \Big[P_i(\theta; \boldsymbol{\gamma}_i)^{x_i} (1 - P_i(\theta; \boldsymbol{\gamma}_i))^{1-x_i} \Big] \phi(\theta; \mu, \sigma) \, \mathrm{d}\theta, \tag{1}$$

where ϕ denotes the density of the normal distribution with mean μ and standard deviation (SD) σ. The distribution parameters of the real-valued latent variable θ—often referred to as a trait, ability, or ability variable—are contained in the vector $\boldsymbol{\delta} = (\mu, \sigma)$. The vector $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_I)$ encompasses the item parameters of item response functions (IRFs) $P_i(\theta; \boldsymbol{\gamma}_i) = \mathsf{P}(X_i = 1 | \theta)$ $(i = 1, \ldots, I)$. The IRF of the two-parameter logistic (2PL) model [4] is defined as

$$P_i(\theta; \boldsymbol{\gamma}_i) = \Psi(a_i(\theta - b_i)) \tag{2}$$

using the item discrimination a_i and item difficulty b_i, and $\Psi(x) = (1 + \exp(-x))^{-1}$ denotes the logistic distribution function. For independently and identically distributed observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ of N cases from the distribution of the random variable $\boldsymbol{X}$, the unknown model parameters of the IRT model (1) can be consistently estimated with a marginal maximum likelihood estimation [5–7]. It is important to note that some identification constraints are typically required because item and distribution parameters cannot be disentangled otherwise [8].

IRT models are commonly employed to compare the distribution of $\boldsymbol{X}$ across two groups, focusing on the parameters associated with the distribution of the ability variable θ in the IRT model (1). This paper examines linking methods [9] for conducting group comparisons. Linking methods involve a two-step process. First, the IRT model is estimated separately for each group. In the second step, differences in the estimated item parameters are utilized to compute distributional differences of the θ variable across the two groups using a linking method [9–14].

In this article, we apply linking methods in the presence of differential item functioning (DIF; [15–17]). DIF implies a heterogeneous functioning of items across groups, that is, item discrimination or item difficulty in the 2PL model could differ across the two groups. It has been shown that the presence of random DIF [18,19] can result in biased distribution parameter estimates in nonlinear linking methods [20].

The Stocking–Lord (SL) linking method [21], as a nonlinear linking method [9], also yields biased distribution parameter estimates in the presence of random DIF effects [20]. To this end, the simulation extrapolation (SIMEX; [22]) measurement error correction method has been applied in SL linking [23]. It turns out that the SIMEX method is effective in reducing the bias in SL linking [23]. However, a purely simulation-based SIMEX method can be computationally demanding requiring repeated application of the iterative SL linking procedure many times. Therefore, two computational shortcuts are introduced and compared with the original SIMEX procedure in this article through a simulation study. It is hypothesized that these shortcuts will yield results comparable to those of the original SIMEX-based SL linking procedure.

The article proceeds with the following structure. Section 2 briefly describes the SL linking method. Section 3 reviews the general concept of the SIMEX method and its application to SL linking. Furthermore, two computational shortcuts to SIMEX-based SL linking are proposed in this section. Section 4 presents the findings of a simulation study comparing the performance of estimators from different SIMEX implementations of SL linking. The article concludes with a discussion in Section 5.

2. Stocking–Lord Linking

In this section, we briefly describe the SL linking approach. Assume that the 2PL model holds in both groups. In the first group, the item discriminations and item intercepts are defined as $a_{i1} = a_i$ and $b_{i1} = b_i$, respectively. For identification purposes, we further assume $\theta \sim \mathsf{N}(0,1)$ in this group. In contrast, for the second group, the ability distribution follows $\theta \sim \mathsf{N}(\mu, \sigma^2)$, where μ and σ are the mean and the SD, respectively. The item discriminations are assumed to be invariant across groups, implying that $a_{i2} = a_i$. Additionally, a random uniform DIF [16] effect e_i is added to the common item difficulty b_i to obtain the item difficulty b_{i2} in the second group. That is, we define

$$b_{i2} = b_i + e_i \text{ , where } \mathsf{E}(e_i) = 0 \text{ and } \mathsf{Var}(e_i) = \tau^2. \tag{3}$$

The variance τ^2 is referred to as the DIF variance [24], and τ is designated as the DIF SD.

We now describe the first step of the linking approach. The 2PL model is separately fitted for each group, assuming a standard normal distribution $\mathsf{N}(0,1)$ for the ability variable θ. The estimated item parameters $\widehat{a}_{i1}$ and $\widehat{b}_{i1}$ will deviate from the true data-generating parameters $a_{i1} = a_i$ and $b_{i1} = b_i$ due to sampling error (of subjects). In the second group, the original item parameters a_{i2} and b_{i2} cannot be directly recovered, as the 2PL model is fit with $\theta \sim \mathsf{N}(0,1)$, while the true model specifies $\theta \sim \mathsf{N}(\mu, \sigma^2)$. Straightforward algebra yields

$$a_{i2}(\theta - b_{i2}) = a_{i2}(\sigma\theta^* + \mu - b_{i2}) \text{ for } \theta^* \sim \mathsf{N}(0,1). \tag{4}$$

Therefore, the identified item parameters in the second group can be expressed as

$$a_{i2}^* = a_i\sigma \text{ and } b_{i2}^* = \sigma^{-1}(b_i + e_i - \mu). \tag{5}$$

Due to sampling errors, the estimated item parameters $\widehat{a}_{i1}$ and $\widehat{b}_{i1}$ (or $\widehat{a}_{i2}$ and $\widehat{b}_{i2}$) will also deviate from a_{i1}^* and b_{i1}^* (or a_{i2}^* and b_{i2}^*).

We now collect all estimated item parameters in vectors $\widehat{\boldsymbol{a}}_g$ and $\widehat{\boldsymbol{b}}_g$ for the two groups $g = 1, 2$. The Stocking–Lord linking method, as introduced in [21], employs the linking function H that is defined as

$$H(\mu, \sigma, \widehat{\boldsymbol{a}}_1, \widehat{\boldsymbol{b}}_1, \widehat{\boldsymbol{a}}_2, \widehat{\boldsymbol{b}}_2) = \sum_{t=1}^{T} \omega_t \left[\sum_{i=1}^{I} \Psi(\widehat{a}_{i1}(\sigma\theta_t + \mu - \widehat{b}_{i1})) - \sum_{i=1}^{I} \Psi(\widehat{a}_{i2}(\theta_t - \widehat{b}_{i2})) \right]^2. \tag{6}$$

The weights ω_t are known, and a grid for the ability variable θ is chosen, ranging between θ_1 and θ_T for θ_t with $t = 1, \ldots, T$. In applications, the grid could be chosen to have equidistant values, and the weights ω_t could either be equal to 1 or are proportional to a discretized density of a normal distribution with an SD σ_0 larger than 1 (e.g., an SD of 2), that is, one chooses $\omega_t \propto \exp\left(-\theta_t^2/(2\sigma_0^2)\right)$ with $\sum_{t=1}^{T} \omega_t = 1$. We would like to point out that the linking function (6) is referred to as asymmetric Stocking–Lord linking because it aligns the test characteristic function (TCF) from the first group with that of the second group.

The distribution parameters μ and σ can be estimated as the minimizer of a general linking function H:

$$(\widehat{\mu}, \widehat{\sigma}) = \underset{(\mu,\sigma)}{\arg\min}\ H(\mu, \sigma, \widehat{\boldsymbol{a}}_1, \widehat{\boldsymbol{b}}_1, \widehat{\boldsymbol{a}}_2, \widehat{\boldsymbol{b}}_2). \tag{7}$$

For a differentiable function H, the estimating equations for μ and σ are obtained by computing the partial derivatives of H with respect to μ and σ. Obviously, the SL linking function defined in (6) is differentiable with respect to μ and σ. Note that the minimization of the SL linking function H requires an iterative estimation, as no closed-form solution for the parameter estimate $(\widehat{\mu}, \widehat{\sigma})$ is available.

3. Proposed Computational Shortcuts in SIMEX-Based Linking

In this section, the SIMEX method is reviewed, and its application to SL linking is explained. Section 3.1 provides an overview of the general principle of SIMEX as a statistical technique. In Section 3.2, the application of SIMEX to linking methods in the presence of random DIF is discussed. Next, Section 3.3 describes the first computational shortcut, the item replication method, for SIMEX-based linking. Finally, Section 3.4 demonstrates how approximate one-step estimates for linking methods can be derived using a Taylor expansion, which is particularly useful for SIMEX-based linking methods that are computationally intensive.

3.1. A Glimpse of the SIMEX Technique

This section outlines the core principle of the SIMEX method introduced by Refs. [22,25,26]. The SIMEX is a general approach for providing bias correction due to measurement errors in variables for any multivariate statistical method. Consider a dataset composed of two parts: a multivariate random variable $\boldsymbol{Z}$ measured without error, and a univariate random variable W, which contains measurement errors. Assume that independent realizations $\boldsymbol{z}_i$ and w_i of $\boldsymbol{Z}$ and W are available for cases $i = 1, \ldots, I$ in a sample of I cases. Furthermore, assume the classical measurement error model [27]

$$w_i = w_i^* + e_i, \tag{8}$$

where w_i^* represents the measured value of W without error, and e_i denotes the variable referring to the measurement error. In the following, we suppose that the measurement error variance $\mathrm{Var}(e_i) = \tau^2$ is either known or can be consistently estimated from the data. Additionally, we make the assumption that the measurement errors e_i follow a normal distribution with a mean of zero.

A statistical procedure yields a parameter estimate $\widehat{\boldsymbol{\delta}} = f(\boldsymbol{Z}, W)$ for a parameter vector $\boldsymbol{\delta}$, where f is a known function of the dataset represented by the random variables $\boldsymbol{Z}$ and W. The known function f can represent any parametric multivariate statistical method. The bias of the estimated parameter $\widehat{\boldsymbol{\delta}}$ is affected by the measurement error variance τ^2. The SIMEX method seeks to eliminate or at least mitigate this bias in $\widehat{\boldsymbol{\delta}}$ caused by measurement errors through a simulation-based technique. The core idea is to intentionally introduce additional measurement errors into the data and then compute the parameter estimate $\widehat{\boldsymbol{\delta}}$ based on these altered datasets that include extra measurement errors. The values w_i of variable W are replaced with

$$\widetilde{w}_i = w_i + \widetilde{e}_i = w_i^* + e_i + \widetilde{e}_i, \tag{9}$$

where $\widetilde{e}_i$ represents a random draw from a normal distribution with a mean of zero and a variance of $\lambda\tau^2$ with $\lambda \geq 0$. As a result, the adjusted values $\widetilde{w}_i$ (i.e., the pseudo-values) contain measurement error $e_i + \widetilde{e}_i$, leading to a variance of $(1+\lambda)\tau^2$. The SIMEX method provides estimates $\widehat{\boldsymbol{\delta}}$ as a function of λ across a grid of λ values. The literature suggests using the following grid of λ values: 0.5, 1.0, 1.5, and 2.0 (see [27,28]). To mitigate Monte Carlo errors due to simulating extra measurement errors in $\widetilde{e}_i$ (see (9)), the estimation is carried out through repeated simulations of the dataset for each fixed λ value. As a result, the parameter estimate is expressed as a function of λ, leading to a parameter curve $\widehat{\boldsymbol{\delta}}(\lambda)$. This first stage of the procedure pertains to the simulation step in SIMEX, while the second stage involves an extrapolation. In that stage, a regression function $\widehat{\boldsymbol{\delta}}(\lambda)$ is estimated, and the predicted value at $\lambda = -1$ yields a bias-reduced estimate for $\boldsymbol{\delta}$ that has been extrapolated to a scenario with no measurement error variance (i.e., $\tau^2 = 0$). Any functional form for $\widehat{\boldsymbol{\delta}}(\lambda)$ could be assumed. However, a quadratic extrapolating function [27] is found to be effective in applications and frequently recommended in the literature using the specification

$$\widehat{\boldsymbol{\delta}}(\lambda) \approx \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1\lambda + \boldsymbol{\alpha}_2\lambda^2. \tag{10}$$

The final parameter estimate of $\boldsymbol{\delta}$ in SIMEX is given by $\widehat{\boldsymbol{\delta}}(-1) = \boldsymbol{\alpha}_0 - \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$. The SIMEX method is a highly adaptable approach for correcting measurement errors, as it is applicable to any class of statistical models [27,29].

In the application of SIMEX to linking methods, $\boldsymbol{Z}$ represents the vector of input item parameters, and W refers to DIF effects. The key idea is that the presence of DIF effects introduces a bias, which can be interpreted as a measurement error in SIMEX. Researchers aim to obtain unbiased linking parameter estimates corresponding to a situation where DIF effects are absent. In this context, DIF effects, viewed as variables prone to measurement errors, should not influence the linking procedure. SIMEX offers a method to eliminate the bias in a linking method caused by measurement errors, specifically the presence of random DIF.

3.2. Applying the SIMEX Approach to Linking Methods

The SIMEX method can be used to reduce the bias in SL linking [23] if uniform DIF is present. The idea is to interpret DIF effects as measurement errors that can be handled with SIMEX. The parameter estimate $\widehat{\delta} = (\widehat{\mu}, \widehat{\sigma})$ in SL linking as obtained by

$$\widehat{\delta} = (\mu^*, \sigma^*) = \arg\min_{\delta} H(\delta, \widehat{a}_1, \widehat{b}_1, \widehat{a}_2, \widehat{b}_2). \tag{11}$$

As shown in Refs. [20,30], the bias of $\widehat{\delta}$ is primarily a function of the DIF variance $\tau^2 = \mathsf{Var}(e_i)$ and can be expressed as

$$\text{Bias}(\widehat{\delta}) = -\frac{1}{2}\tau^2 \boldsymbol{H}_{\delta\delta}(\delta_0; \boldsymbol{\gamma}, \mathbf{0})^{-1} \sum_{i=1}^{I} \boldsymbol{H}_{\delta e_i e_i}(\delta_0; \boldsymbol{\gamma}, \mathbf{0}), \tag{12}$$

where $\boldsymbol{H}_{\delta\delta}(\delta_0; \boldsymbol{\gamma}, \mathbf{0})$ denotes the Hessian matrix of the linking function H with respect to δ, evaluated at the true linking parameter δ_0, joint item parameters $\boldsymbol{\gamma}$ and no DIF effects (i.e., $\boldsymbol{e} = (e_1, \ldots, e_I) = \mathbf{0}$). Moreover, the random DIF variance τ^2 is item-wise weighted by $\boldsymbol{H}_{\delta e_i e_i}(\delta_0; \boldsymbol{\gamma}, \mathbf{0})$, where $\boldsymbol{H}_{\delta e_i e_i}$ involves second-order derivatives of the gradient $\boldsymbol{H}_{\delta}$ with respect to DIF effects e_i, evaluated at $\boldsymbol{e} = \mathbf{0}$.

The application of the SIMEX method to a linking method requires an estimate of the DIF variance τ^2. The true uniform DIF effects e_i are unknown, but can be estimated using the formula

$$\widehat{e}_i = \widehat{\mu} + \widehat{\sigma}\widehat{b}_{i2} - \widehat{b}_{i1}. \tag{13}$$

Subsequently, the DIF variance τ^2 can be consistently estimated by

$$\hat{\tau}^2 = \frac{1}{I}\sum_{i=1}^{I}(\widehat{e}_i - \bar{e})^2 - \frac{1}{I}\sum_{i=1}^{I} v_{\widehat{e}_i}, \tag{14}$$

where $v_{\widehat{e}_i}$ represents the variance estimate due to sampling error of the estimated DIF effect estimate $\widehat{e}_i$ and $\bar{e} = I^{-1}\sum_{i=1}^{I}\widehat{e}_i$.

In the application of SIMEX to SL linking, adjusted DIF effects $e_i^*(\lambda) = \widehat{e}_i + u_i(\lambda)$ are calculated for $\lambda = 0.5$, 1.0, 1.5, and 2.0, where the values $u_i(\lambda)$ are random draws from a normal distribution with zero mean and a variance $\lambda\hat{\tau}^2$. The newly created DIF effects $e_i^*(\lambda)$ have a larger variance of $(1+\lambda)\hat{\tau}^2$. Next, adjusted item difficulties $b_{i2}^*(\lambda)$ (i.e., pseudo-item parameters) are computed as

$$b_{i2}^*(\lambda) = \frac{1}{\sigma^*}\left(\widehat{b}_{i1} + e_i^*(\lambda) - \mu^*\right) = \widehat{b}_{i2} + \frac{1}{\sigma^*}u_i(\lambda). \tag{15}$$

All pseudo-item parameters $b_{i2}^*(\lambda)$ are then gathered into the vector $\boldsymbol{b}_2^*(\lambda)$. The SIMEX-based linking estimate of the mean μ and the standard deviation σ are now a function of λ and are defined as

$$\widehat{\delta}(\lambda) = (\widehat{\mu}(\lambda), \widehat{\sigma}(\lambda)) = \arg\min_{\delta} H(\delta, \widehat{a}_1, \widehat{b}_1, \widehat{a}_2, \boldsymbol{b}_2^*(\lambda)). \tag{16}$$

These estimates are derived from item parameters that exhibit a larger DIF variance than those present in the original data. In the second stage of SIMEX, the regression functions $\widehat{\mu}(\lambda)$ and $\widehat{\sigma}(\lambda)$ are specified as a quadratic function of λ (see (10)). The SIMEX-based estimates correspond to the extrapolated values of the regression functions at $\lambda = -1$; that is, $\widehat{\mu}(-1)$ and $\widehat{\sigma}(-1)$.

The SIMEX method, in its original proposal, involves simulating new datasets. To reduce Monte Carlo estimation errors, simulations are performed multiple times for a fixed λ

value, and the resulting parameter estimates are averaged. In the context of SL linking, a significant challenge arises from the limited number of cases (i.e., items), which results in relatively large Monte Carlo errors. Notably, reducing the Monte Carlo errors necessitates a substantial number of simulations. Therefore, two computational shortcuts are discussed in Sections 3.3 and 3.4, that circumvent the computationally demanding repeated estimations of SL linking in the Monte Carlo approach.

3.3. Replication Method to SIMEX-Based Linking

In the so-called replication approach to SIMEX estimation [23], the simulation step is replaced by a deterministic step in which a pseudo-dataset is generated that resembles the distribution of pseudo-item parameters at the level of the population.

To this end, a quasi-Monte Carlo variant of SIMEX is employed, where SL linking is applied to a created dataset containing MI pseudo-items for an integer M. Note that the original dataset contains I items. The approach involves duplicating item parameters and systematically introducing normally distributed uniform DIF across all items. Let $\varepsilon_1, \ldots, \varepsilon_M$ approximately normally distributed that are generated by a quasi-random Halton sequence generator [31] as implemented in the R function `sirt::qmc.nodes()` [32], which depends on the implementation in `sfsmisc::QUnif()` [33] for the generation of uniform quasi-random numbers. For a fixed λ, the values $\sqrt{\lambda}\tau\varepsilon_m$ have approximately a zero mean and a variance $\lambda\tau^2$. The item difficulty for the pseudo-item (i, m) for $i = 1, \ldots, I$ and $m = 1, \ldots, M$ is defined as

$$\widetilde{b}_{(i,m),2} = \widehat{\sigma}^{-1}(\widehat{b}_{i1} + \widehat{e}_i + \sqrt{\lambda}\tau\varepsilon_m - \widehat{\mu}). \tag{17}$$

Using that method, a set of item parameters for pseudo-items is created, resulting in a DIF variance of $\lambda\tau^2$. All other item parameters $\widehat{a}_{i1}$, $\widehat{b}_{i1}$, and $\widehat{a}_{i2}$ are duplicated for the values of $m = 1, \ldots, M$. For example, the item difficulties in the first group are given by $\widetilde{b}_{(i,m),1} = \widehat{b}_{i1}$ for all $m = 1, \ldots, M$. The SL linking method is utilized on the set of pseudo-items for different λ values, yielding a SIMEX parameter curve $\boldsymbol{\delta}(\lambda)$. It is important to note that our modified SIMEX method eliminates Monte Carlo simulation errors but at the price of only removing (or reducing) asymptotic bias. This means that the SIMEX parameter estimate is essentially evaluated in the case of a large number of items because MI is typically large compared to the original number of items I. As in the simulation-based SIMEX approach, the extrapolation of the fitted quadratic regression functions to the values $\boldsymbol{\delta}(\lambda)$ at $\lambda = -1$ provides a final SIMEX-based linking parameter estimate for $\boldsymbol{\delta}$.

3.4. Approximate Estimation in SIMEX-Based Linking

In this section, we describe a computational shortcut to SIMEX-based linking. The idea is to avoid the computationally demanding iterative estimation of SL linking and replace the estimation with a one-step updating approach that is based on a Taylor approximation of the linking function. The technique has also been used to derive the bias and variance of nonlinear linking functions [20,30].

Let $\widehat{\boldsymbol{\delta}} = (\widehat{\mu}, \widehat{\sigma})$ represent the vector of parameter estimates that is the root of

$$H_{\delta}(\widehat{\boldsymbol{\delta}}, \boldsymbol{a}_1, \boldsymbol{b}_1, \boldsymbol{a}_2, \boldsymbol{b}_2) = \mathbf{0}, \tag{18}$$

where H_{δ} is the vector of partial derivatives of the linking function H with respect to the parameter vector $\boldsymbol{\delta} = (\mu, \sigma)$. It is vital to highlight that the notation in (18) refers to original item parameters $\boldsymbol{a}_g$ and $\boldsymbol{b}_g$ rather than the identified item parameters $\widehat{\boldsymbol{a}}_g$ and $\widehat{\boldsymbol{b}}_g$ for $g = 1, 2$. In the SIMEX-based approach to linking, we substitute the item parameters $\boldsymbol{b}_2 = (b_{12}, \ldots, b_{I2})$ with pseudo-item parameters $\boldsymbol{b}_2^* = (b_{12}^*, \ldots, b_{I2}^*)$ that include an addi-

tional error variance, depending on the parameter λ. The parameter vector $\boldsymbol{\delta}^*$ based on pseudo-item parameters in SIMEX is obtained as the root of

$$H_{\delta}(\boldsymbol{\delta}^*, \boldsymbol{a}_1, \boldsymbol{b}_1, \boldsymbol{a}_2, \boldsymbol{b}_2^*) = \mathbf{0}. \tag{19}$$

In the following, we apply a Taylor expansion of H_{δ} in (19) around $\widehat{\boldsymbol{\delta}}$ in the first order and $\boldsymbol{b}_2$ up to the second order. We obtain

$$\begin{aligned} \mathbf{0} &= H_{\delta}(\boldsymbol{\delta}^*, \boldsymbol{a}, \boldsymbol{b}_1, \boldsymbol{a}, \boldsymbol{b}_2^*) \\ &= H_{\delta}(\widehat{\boldsymbol{\delta}}, \boldsymbol{a}_1, \boldsymbol{b}_1, \boldsymbol{a}_2, \boldsymbol{b}_2) + H_{\delta\delta}(\widehat{\boldsymbol{\delta}}, \boldsymbol{a}_1, \boldsymbol{b}_1, \boldsymbol{a}_2, \boldsymbol{b}_2)\left(\boldsymbol{\delta}^* - \widehat{\boldsymbol{\delta}}\right) \\ &\quad + \sum_{i=1}^{I} H_{\delta b_{i2}}(\widehat{\boldsymbol{\delta}}, \boldsymbol{a}_1, \boldsymbol{b}_1, \boldsymbol{a}_2, \boldsymbol{b}_2)(b_{i2}^* - b_{i2}) \\ &\quad + \frac{1}{2}\sum_{i=1}^{I} H_{\delta b_{i2} b_{i2}}(\widehat{\boldsymbol{\delta}}, \boldsymbol{a}_1, \boldsymbol{b}_1, \boldsymbol{a}_2, \boldsymbol{b}_2)(b_{i2}^* - b_{i2})^2. \end{aligned} \tag{20}$$

Due to (18), we obtain an approximate estimate in a one-step iteration as

$$\boldsymbol{\delta}^* - \widehat{\boldsymbol{\delta}} = -H_{\delta\delta}^{-1}\left[\sum_{i=1}^{I} H_{\delta b_{i2}}(b_{i2}^* - b_{i2}) + \frac{1}{2}\sum_{i=1}^{I} H_{\delta b_{i2} b_{i2}}(b_{i2}^* - b_{i2})^2\right]. \tag{21}$$

Arguments in the derivatives are omitted in (21). For instance, we use the short notation $H_{\delta\delta} = H_{\delta\delta}(\widehat{\boldsymbol{\delta}}, \boldsymbol{a}_1, \boldsymbol{b}_1, \boldsymbol{a}_2, \boldsymbol{b}_2)$. As a consequence, the approximation (21) to the linking estimation problem is a linear function in $\boldsymbol{\delta}$. Moreover, the dependence with respect to $\boldsymbol{b}_2$ is allowed to be quadratically related, and the approximation error will typically decrease for smaller differences $\boldsymbol{b}_2^* - \boldsymbol{b}_2$. The approximate estimation of the linking method can be utilized in both cases, for pseudo-item parameters $\boldsymbol{b}_2^*$ that emerge from simulation or replication. However, the reduction in computational demand is more important for SIMEX-based linking that relies on simulation.

In an implementation of an approximate estimation in SIMEX-based SL linking, the required partial derivatives can be analytically computed in a straightforward way or can be computed via numerical differentiation.

4. Simulation Study

4.1. Method

This simulation study employed the 2PL model to generate item responses of subjects from two groups. In the first group, the ability variable θ in the IRT model defined in (1) was normally distributed with a mean of zero and a standard deviation of one. For the second group, the ability variable θ had a mean of $\mu = 0.3$ and a standard deviation of $\sigma = 1.2$.

The number of items I in the simulation was set to 10, 20, or 30. The group-specific item parameters a_{ig} and b_{ig} in the 2PL model for $i = 1, \ldots, I$ and $g = 1, 2$ were based on common item parameters a_i and b_i, which remained fixed across replications in the simulation, along with randomly generated uniform and normally distributed DIF effects. Note that uniform DIF effects e_i were simulated in each replication of the simulation study.

For the case of $I = 10$ items, the common item discriminations a_i were selected as follows: 1.14, 1.05, 1.13, 0.79, 0.70, 1.13, 1.28, 0.87, 0.80, and 1.11. The mean of the a_i parameters was $M = 1.000$ and the standard deviation was $SD = 0.194$, representing a test with moderately discriminating items and moderate variability in item discrimination. The common item difficulties b_i were chosen as 0.86, -0.60, 1.45, -1.18, -0.71, -0.94,

-0.19, 0.17, 1.55, and -0.41. The b_i parameters had a mean $M = 0.000$ and a standard deviation $SD = 0.979$, resulting in a test with items of average difficulty. For item counts of $I = 20$ and $I = 30$, the item parameters of the 10 items were duplicated in multiples of 10.

Uniform DIF effects were simulated by adding a normally distributed random DIF effect with zero mean and DIF variance τ^2 to item difficulties in the second group. More formally, we defined $b_{i1} = b_i$ and $b_{i2} = b_i + e_i$. The group-specific item discriminations a_{ig} were chosen to be equal (i.e., $a_{i1} = a_{i2} = a_i$). The DIF SD τ was chosen as 0, 0.25, and 0.50, referring to situations of no DIF, small DIF, and large DIF.

Furthermore, item responses from the 2PL model were simulated for finite sample sizes $N = 500$, 1000, 2000, and 4000.

In each of the 4 (sample size N) $\times$ 3 (number of items I)$\times$ 3 (DIF SD τ) = 36 cells of the simulation, 3000 replications were conducted.

In total, five different variants of SL linking were applied to each of the simulated datasets. First, method "SL" denotes the original SL linking that is prone to bias in the presence of random DIF. The SL linking methods "RE" and "RA" denote the SIMEX-based linking approach utilizing replication (see Section 3.3) and the exact and approximate (see Section 3.4) SL estimate, respectively. The replication methods in this simulation study used a replication factor of $M = 41$ (see Section 3.3). For example, the pseudo-datasets in SIMEX included $MI = 820$ pseudo-items in the condition with $I = 20$ items. The SL methods "SE" and "SA" denote the simulation-based linking methods utilizing the exact and approximate SL estimates, respectively. The simulation approach for SIMEX in this study used 100 Monte Carlo replications for the computation of the parameter curves $\widehat{\boldsymbol{\delta}}(\lambda)$.

For all five SL methods, the empirical bias was computed for the estimated group mean $\widehat{\mu}$ and the estimated SD $\widehat{\sigma}$. Let ψ denote the parameter μ or σ and let $\widehat{\psi}$ represent the corresponding linking parameter estimate. Additionally, let $\widehat{\psi}_r$ denote the parameter estimate in the rth replicated dataset for $r = 1, \ldots, R$. The empirical bias was given by

$$\text{Bias}(\widehat{\psi}) = \frac{1}{R}\sum_{r=1}^{R}\widehat{\psi}_r - \psi. \tag{22}$$

The empirical root mean square error (RMSE) of the estimates was calculated as

$$\text{RMSE}(\widehat{\psi}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\widehat{\psi}_r - \psi)^2}. \tag{23}$$

The relative RMSE was calculated by dividing the RMSE of a particular method by the RMSE of the simulation-based SIMEX method with approximate estimation (i.e., method SA), and then multiplying the result by 100. Relative RMSE values smaller than 100 indicated methods that outperformed method SA in terms of precision.

The analysis of this simulation study was conducted with the statistical software R (Version 4.4.1; [34]). The 2PL model was fitted with the `sirt::xxirt()` function from the R package sirt (Version 4.2-73; [32]). The author of this article developed custom R functions for the linking methods investigated in this simulation. These functions, along with the replication material for this Simulation Study, can be accessed at https://osf.io/b4rdh (accessed on 3 October 2024).

4.2. Results

Table 1 presents bias of the estimated mean $\widehat{\mu}$ and the estimated SD $\widehat{\sigma}$ as a function of the DIF SD τ, the number of items I, and the sample size N. All linking methods were unbiased under the no DIF conditions. In the small DIF condition (i.e, $\tau = 0.25$), the bias was negligible. In the large DIF condition (i.e., $\tau = 0.5$), the mean $\widehat{\mu}$ was slightly biased,

and the SD $\widehat{\sigma}$ was noticeably biased for the original SL linking method. Interestingly, the simulation-based SIMEX methods SE and SA had a bias in the case of a small number of items (i.e., $I = 10$). This bias vanished for longer tests with a larger number of items. The replication-based methods RE and RA had the least bias.

Table 1. Simulation Study: bias of the estimated mean $\widehat{\mu}$ and the estimated standard deviation $\widehat{\sigma}$ as a function of the uniform DIF standard deviation τ, number of items I, and sample size per group N.

		$\widehat{\mu}$					$\widehat{\sigma}$				
I	***N***	**SL**	**RE**	**RA**	**SE**	**SA**	**SL**	**RE**	**RA**	**SE**	**SA**
		No DIF (DIF SD $\tau = 0$)									
10	500	0.005	0.005	0.005	0.004	0.004	0.007	0.008	0.008	0.008	0.008
	1000	0.001	0.001	0.001	0.000	0.000	0.003	0.004	0.004	0.004	0.004
	2000	−0.001	−0.001	−0.001	−0.001	−0.001	0.002	0.002	0.002	0.002	0.002
	4000	0.000	0.000	0.000	0.000	0.000	−0.001	−0.001	−0.001	−0.001	−0.001
20	500	0.002	0.002	0.002	0.002	0.002	0.005	0.006	0.006	0.006	0.006
	1000	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
	2000	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.002
	4000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001
30	500	−0.001	−0.001	−0.001	−0.002	−0.002	0.007	0.008	0.008	0.008	0.008
	1000	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.002	0.002	0.002
	2000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.002	0.002	0.002
	4000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001
		Small DIF (DIF SD $\tau = 0.25$)									
10	500	0.001	0.003	0.003	−0.004	−0.004	−0.002	0.006	0.006	0.004	0.004
	1000	0.002	0.004	0.004	−0.004	−0.004	−0.005	0.003	0.003	0.001	0.001
	2000	−0.001	0.001	0.001	−0.006	−0.007	−0.008	0.000	0.000	−0.001	−0.001
	4000	−0.001	0.001	0.001	−0.006	−0.007	−0.008	0.000	0.000	−0.001	−0.001
20	500	−0.001	0.001	0.001	0.003	0.003	−0.004	0.004	0.004	0.004	0.004
	1000	−0.001	0.001	0.001	0.002	0.003	−0.006	0.003	0.003	0.003	0.003
	2000	−0.001	0.001	0.001	0.003	0.003	−0.009	0.001	0.001	0.000	0.000
	4000	−0.003	0.000	0.000	0.001	0.001	−0.009	0.000	0.000	0.000	0.000
30	500	−0.003	0.000	0.000	−0.002	−0.002	−0.007	0.002	0.002	0.002	0.002
	1000	−0.003	−0.001	−0.001	−0.002	−0.002	−0.009	0.001	0.001	0.000	0.000
	2000	−0.003	0.000	0.000	−0.002	−0.002	−0.009	0.001	0.001	0.001	0.001
	4000	−0.004	−0.002	−0.002	−0.003	−0.003	−0.009	0.001	0.001	0.000	0.000
		Large DIF (DIF SD $\tau = 0.5$)									
10	500	−0.005	0.003	0.003	**−0.012**	**−0.013**	**−0.027**	0.005	0.005	−0.001	0.000
	1000	−0.007	0.001	0.001	**−0.014**	**−0.015**	**−0.031**	0.001	0.001	−0.005	−0.004
	2000	−0.007	0.001	0.001	**−0.015**	**−0.015**	**−0.033**	−0.001	−0.001	−0.007	−0.006
	4000	**−0.010**	−0.002	−0.002	**−0.018**	**−0.019**	**−0.035**	−0.002	−0.003	−0.009	−0.008
20	500	−0.008	0.001	0.000	0.002	0.003	**−0.036**	0.000	−0.001	−0.003	−0.003
	1000	**−0.011**	−0.002	−0.003	−0.001	0.000	**−0.036**	−0.001	−0.002	−0.003	−0.003
	2000	**−0.010**	−0.001	−0.001	0.001	0.001	**−0.037**	−0.001	−0.002	−0.004	−0.004
	4000	−0.009	0.000	0.000	0.001	0.002	**−0.037**	−0.001	−0.002	−0.004	−0.004
30	500	**−0.011**	−0.002	−0.002	−0.004	−0.005	**−0.034**	0.003	0.002	0.002	0.001
	1000	**−0.010**	0.000	−0.001	−0.003	−0.003	**−0.034**	0.002	0.001	0.001	0.001
	2000	**−0.010**	0.000	−0.001	−0.003	−0.004	**−0.039**	−0.002	−0.003	−0.004	−0.004
	4000	**−0.011**	−0.002	−0.002	−0.005	−0.005	**−0.038**	−0.001	−0.002	−0.003	−0.003

Note. SL = original Stocking–Lord linking; RE = SIMEX with replication (see Section 3.3), exact estimation; RA = SIMEX with replication (see Section 3.3), approximate estimation (see Section 3.4); SE = SIMEX with simulation, exact estimation; SA = SIMEX with simulation, approximate estimation (see Section 3.4); biases with absolute values larger than 0.010 are printed in bold font.

Table 2 presents the relative RMSE for the estimated parameters $\widehat{\mu}$ and $\widehat{\sigma}$. It is important to emphasize that all four SIMEX-based SL linking methods RE, RA, SE, and SA provided similar RMSE values. Hence, SIMEX-based linking with replication (methods RE and RA), as well as simulation-based linking with approximate estimation (method SA), serve as successful alternatives to the computationally demanding SIMEX-based linking with simulation and exact estimation (i.e., method SE).

Table 2. Simulation Study: relative root mean square error (RMSE) of the estimated mean $\widehat{\mu}$ and the estimated standard deviation $\widehat{\sigma}$ as a function of the uniform DIF standard deviation τ, number of items I, and sample size per group N.

		$\widehat{\mu}$					$\widehat{\sigma}$				
I	N	SL	RE	RA	SE	SA	SL	RE	RA	SE	SA
		No DIF (DIF SD $\tau = 0$)									
10	500	100.0	100.0	100.0	100.0	100	*99.7*	100.0	100.0	100.0	100
	1000	100.0	100.0	100.0	100.0	100	*99.9*	100.0	100.0	100.0	100
	2000	100.0	100.0	100.0	100.0	100	*99.9*	100.0	100.0	100.0	100
	4000	100.0	100.0	100.0	100.0	100	100.0	100.0	100.0	100.0	100
20	500	*99.9*	100.0	100.0	100.0	100	*99.7*	100.0	100.0	100.0	100
	1000	*99.9*	100.0	100.0	100.0	100	*99.9*	100.0	100.0	100.0	100
	2000	100.0	100.0	100.0	100.0	100	100.0	100.0	100.0	100.0	100
	4000	100.0	100.0	100.0	100.0	100	100.0	100.0	100.0	100.0	100
30	500	100.0	100.0	100.0	100.0	100	*99.8*	100.0	100.0	100.0	100
	1000	100.0	100.0	100.0	100.0	100	*99.9*	100.0	100.0	100.0	100
	2000	100.0	100.0	100.0	100.0	100	100.0	100.0	100.0	100.0	100
	4000	100.0	100.0	100.0	100.0	100	100.0	100.0	100.0	100.0	100
		Small DIF (DIF SD $\tau = 0.25$)									
10	500	*99.4*	*99.9*	*99.9*	100.0	100	*99.3*	100.3	100.2	100.0	100
	1000	*99.4*	100.1	100.1	100.0	100	*99.4*	100.3	100.3	100.0	100
	2000	*99.2*	*99.8*	*99.8*	100.0	100	100.4	100.3	100.2	100.1	100
	4000	*99.1*	*99.8*	*99.8*	*99.9*	100	100.8	100.3	100.3	100.1	100
20	500	*99.4*	100.0	100.0	100.0	100	*99.4*	100.1	100.1	100.0	100
	1000	*99.3*	100.0	100.0	100.0	100	*99.5*	100.1	100.1	100.0	100
	2000	*99.2*	100.0	100.0	100.0	100	101.1	100.1	100.1	100.0	100
	4000	*99.4*	100.0	100.0	100.0	100	**102.3**	100.1	100.0	100.0	100
30	500	*99.5*	100.0	100.0	100.0	100	*99.4*	100.1	100.1	100.0	100
	1000	*99.5*	100.0	100.0	100.0	100	100.5	100.0	100.0	100.0	100
	2000	*99.4*	100.0	100.0	100.0	100	101.6	100.1	100.1	100.0	100
	4000	*99.4*	*99.9*	*99.9*	100.0	100	**103.5**	100.1	100.1	100.1	100
		Large DIF (DIF SD $\tau = 0.5$)									
10	500	*97.8*	100.3	100.2	*99.9*	100	*99.8*	101.0	100.9	100.0	100
	1000	*97.6*	100.1	100.1	*99.9*	100	**102.4**	100.8	100.6	100.0	100
	2000	*97.4*	100.0	*99.9*	*99.9*	100	**104.7**	101.0	100.8	100.3	100
	4000	*97.2*	*99.8*	*99.8*	*99.8*	100	**108.5**	100.8	100.6	100.6	100
20	500	*97.8*	100.2	100.1	100.1	100	**105.2**	100.4	100.2	100.1	100
	1000	*97.9*	100.2	100.1	100.0	100	**111.1**	100.3	100.1	100.2	100
	2000	*97.6*	100.2	100.2	100.0	100	**117.5**	100.4	100.3	100.4	100
	4000	*97.5*	100.2	100.2	100.0	100	**125.0**	100.3	100.1	100.3	100
30	500	*98.0*	100.0	100.0	100.0	100	**105.6**	100.3	100.1	100.1	100
	1000	*97.7*	100.1	100.0	100.0	100	**111.9**	100.3	100.0	100.1	100
	2000	*97.6*	100.1	100.0	100.0	100	**126.3**	100.1	100.0	100.2	100
	4000	*97.7*	100.0	*99.9*	100.0	100	**132.4**	100.2	100.0	100.3	100

Note. SL = original Stocking–Lord linking; RE = SIMEX with replication (see Section 3.3), exact estimation; RA = SIMEX with replication (see Section 3.3), approximate estimation (see Section 3.4); SE = SIMEX with simulation, exact estimation; SA = SIMEX with simulation, approximate estimation (see Section 3.4); relative RMSE values larger than 102.0 are printed in bold font. Cells with RMSE values smaller than 100.0 are printed in italic font. The method SA served as the reference method in the computation of the relative RMSE.

5. Discussion

The literature highlights that SL linking produces biased estimates when random DIF is present in item difficulties under the 2PL model. To address this issue, a bias reduction method for SL linking using the SIMEX technique has been proposed. The core idea of applying SIMEX to SL linking is to treat random DIF as covariates affected by measurement

errors. However, SIMEX-based linking is computationally intensive. This is because the nonlinear optimization of the SL linking function requires an iterative estimation, as no closed-form solution exists. Additionally, the SIMEX technique relies on simulation by introducing artificially generated random DIF into the item parameters, necessitating repeated estimation across numerous resimulated datasets. To mitigate the computational burden, this article explored two shortcuts for the SIMEX-based SL linking method, which depends heavily on Monte Carlo simulation. The first shortcut, referred to as the replication approach, substituted the assumed normal distribution of random DIF with a deterministic replication of items. This replication generated DIF items that approximated the exact normal distribution. The second shortcut replaced the iterative estimation in SL linking with a one-step minimization based on a Taylor expansion of the SL linking function. Since the resimulated datasets in SIMEX only slightly perturb the original item parameters, the one-step approach may provide a reasonable approximation.

The two shortcuts were compared to the original SIMEX-based SL linking procedure through a simulation study. Results showed that the repeated iterative estimation in the SL linking method was unnecessary, as the precision of the linking parameter estimation remained essentially unaffected. Consequently, the one-step update of SL linking based on a Taylor approximation proved useful for practical implementation. Furthermore, the one-step update could also be applied to the replication-based approach, which demonstrated superior performance compared to the simulation-based SIMEX SL linking approach when dealing with a small set of ten items. These findings suggest that practitioners can use SIMEX-based SL linking without resorting to computationally intensive routines.

As in any simulation study, our simulation study had several limitations. First, we only handled uniform DIF effects. Although nonuniform DIF effects (i.e., DIF in item discriminations) seem to appear in practical applications less frequently [35], an extension of the proposed SIMEX-based SL linking methods to nonuniform DIF effects might be an interesting topic for future research. Second, SIMEX-based SL linking [36,37] could also be explored for polytomous item responses in the generalized partial credit model [38]. Third, alternative IRT models might be studied in SIMEX-based SL linking, such as the Rasch model or the three- or four-parameter logistic IRT models. Moreover, SIMEX-based SL linking can be investigated in longitudinal linking studies [39]. In addition, the proposed shortcut based on Taylor approximation in the SIMEX method could also be applied in fixed item parameter calibration in future research [40].

Supplementary Materials: The following supporting information can be downloaded at: https://osf.io/b4rdh (accessed on 3 October 2024).

Author Contributions: Conceptualization, A.R.; methodology, A.R.; software, A.R.; validation, A.R.; formal analysis, A.R.; investigation, A.R.; resources, A.R.; data curation, A.R.; writing—original draft preparation, A.R.; writing—review and editing, A.R.; visualization, A.R.; supervision, A.R.; project administration, A.R.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This article only uses simulated datasets. Replication material for creating the simulated datasets in Section 3 can be found at https://osf.io/b4rdh (accessed on 3 October 2024) (Supplementary Materials).

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

2PL	Two-parameter logistic
DIF	Differential item functioning
IRF	Item response function
IRT	Item response theory
RMSE	Root-mean-square error
SD	Standard deviation
SIMEX	Simulation extrapolation
SL	Stocking–Lord
TCF	Test characteristic function

References

1. Chen, Y.; Li, X.; Liu, J.; Ying, Z. Item response theory—A statistical framework for educational and psychological measurement. *Stat. Sci.* 2024, *Epub ahead of print*. Available online: https://rb.gy/1yic0e (accessed on 3 October 2024).
2. Bock, R.D.; Moustaki, I. Item response theory in a general framework. In *Handbook of Statistics, Volume 26: Psychometrics*; Rao, C.R., Sinharay, S., Eds.; Elservier: Amsterdam, The Netherlands, 2007; pp. 469–513. [CrossRef]
3. van der Linden, W.J. Unidimensional logistic response models. In *Handbook of Item Response Theory, Volume 1: Models*; van der Linden, W.J., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 11–30. [CrossRef]
4. Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores*; Lord, F.M., Novick, M.R., Eds.; MIT Press: Reading, MA, USA, 1968; pp. 397–479.
5. Aitkin, M. Expectation maximization algorithm and extensions. In *Handbook of Item Response Theory, Volume 2: Statistical Tools*; van der Linden, W.J., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 217–236. [CrossRef]
6. Bock, R.D.; Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **1981**, *46*, 443–459. [CrossRef]
7. Glas, C.A.W. Maximum-likelihood estimation. In *Handbook of Item Response Theory, Volume 2: Statistical Tools*; van der Linden, W.J., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 197–216. [CrossRef]
8. San Martin, E. Identification of item response theory models. In *Handbook of Item Response Theory, Volume 2: Statistical Tools*; van der Linden, W.J., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 127–150. [CrossRef]
9. Kolen, M.J.; Brennan, R.L. *Test Equating, Scaling, and Linking*; Springer: New York, NY, USA, 2014. [CrossRef]
10. Battauz, M. *equateMultiple: Equating of Multiple Forms*, R Package Version 1.0.0; 2024. Available online: https://cran.r-project.org/web/packages/equateMultiple/index.html (accessed on 13 September 2024). [CrossRef]
11. Battauz, M. Multiple equating of separate IRT calibrations. *Psychometrika* **2017**, *82*, 610–636. [CrossRef] [PubMed]
12. González, J.; Wiberg, M. *Applying Test Equating Methods. Using R*; Springer: New York, NY, USA, 2017. [CrossRef]
13. Lee, W.C.; Lee, G. IRT linking and equating. In *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test*; Irwing, P., Booth, T., Hughes, D.J., Eds.; Wiley: New York, NY, USA, 2018; pp. 639–673. [CrossRef]
14. Sansivieri, V.; Wiberg, M.; Matteucci, M. A review of test equating methods with a special focus on IRT-based approaches. *Statistica* **2017**, *77*, 329–352. [CrossRef]
15. Holland, P.W.; Wainer, H. (Eds.) *Differential Item Functioning: Theory and Practice*; Routledge: London, UK, 1993. [CrossRef]
16. Millsap, R.E. *Statistical Approaches to Measurement Invariance*; Routledge: New York, NY, USA, 2011. [CrossRef]
17. Penfield, R.D.; Camilli, G. Differential item functioning and item bias. In *Handbook of Statistics, Volume 26: Psychometrics*; Rao, C.R., Sinharay, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2007; pp. 125–167. [CrossRef]
18. De Boeck, P. Random item IRT models. *Psychometrika* **2008**, *73*, 533–559. [CrossRef]
19. Fox, J.P.; Verhagen, A.J. Random item effects modeling for cross-national survey data. In *Cross-Cultural Analysis: Methods and Applications*; Davidov, E., Schmidt, P., Billiet, J., Eds.; Routledge: London, UK, 2010; pp. 461–482. [CrossRef]
20. Robitzsch, A. Bias-reduced Haebara and Stocking-Lord linking. *J* **2024**, *7*, 373–384. [CrossRef]
21. Stocking, M.L.; Lord, F.M. Developing a common metric in item response theory. *Appl. Psychol. Meas.* **1983**, *7*, 201–210. [CrossRef]
22. Carroll, R.J.; Küchenhoff, H.; Lombard, F.; Stefanski, L.A. Asymptotics for the SIMEX estimator in nonlinear measurement error models. *J. Am. Stat. Assoc.* **1996**, *91*, 242–250. [CrossRef]
23. Robitzsch, A. SIMEX-based and analytical bias corrections in Stocking-Lord linking. *Analytics* **2024**, *3*, 368–388. [CrossRef]
24. Longford, N.T.; Holland, P.W.; Thayer, D.T. Stability of the MH D-DIF statistics across populations. In *Differential Item Functioning*; Holland, P.W., Wainer, H., Eds.; Routledge: London, UK, 1993; pp. 171–196.
25. Cook, J.R.; Stefanski, L.A. Simulation-extrapolation estimation in parametric measurement error models. *J. Am. Stat. Assoc.* **1994**, *89*, 1314–1328. [CrossRef]

26. Stefanski, L.A.; Cook, J.R. Simulation-extrapolation: The measurement error jackknife. *J. Am. Stat. Assoc.* **1995**, *90*, 1247–1256. [CrossRef]
27. Carroll, R.J.; Ruppert, D.; Stefanski, L.A.; Crainiceanu, C.M. *Measurement Error in Nonlinear Models: A Modern Perspective*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2006. [CrossRef]
28. Lederer, W.; Küchenhoff, H. A short introduction to the SIMEX and MCSIMEX. *R News* **2006**, *6*, 26–31. Available online: https://journal.r-project.org/articles/RN-2006-031/ (accessed on 3 October 2024).
29. Buonaccorsi, J.P. *Measurement Error: Models, Methods, and Applications*; CRC Press: Boca Raton, FL, USA, 2010. [CrossRef]
30. Robitzsch, A. Does random differential item functioning occur in one or two groups? Implications for bias and variance in asymmetric and symmetric Haebara and Stocking-Lord linking. *Asymmetry* **2024**, *1*, 0005. [CrossRef]
31. Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods*; SIAM: Philadelphia, PA, USA, 1992. [CrossRef]
32. Robitzsch, A. *sirt: Supplementary Item Response Theory Models*, R Package Version 4.2-73; 2024. Available online: https://github.com/alexanderrobitzsch/sirt (accessed on 7 September 2024).
33. Maechler, M. *sfsmisc: Utilities from 'Seminar fuer Statistik' ETH Zurich*, R package version 1.1-19; 2024. Available online: https://cran.r-project.org/web/packages/sfsmisc/index.html (accessed on 13 September 2024). [CrossRef]
34. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2024. Available online: https://www.R-project.org (accessed on 15 June 2024).
35. Rutkowski, L.; Svetina, D. Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educ. Psychol. Meas.* **2014**, *74*, 31–57. [CrossRef]
36. Andersson, B. Asymptotic variance of linking coefficient estimators for polytomous IRT models. *Appl. Psychol. Meas.* **2018**, *42*, 192–205. [CrossRef] [PubMed]
37. Zhang, Z. Asymptotic standard errors of generalized partial credit model true score equating using characteristic curve methods. *Appl. Psychol. Meas.* **2021**, *45*, 331–345. [CrossRef] [PubMed]
38. Muraki, E. A generalized partial credit model: Application of an EM algorithm. *Appl. Psychol. Meas.* **1992**, *16*, 159–176. [CrossRef]
39. Engels, O.; Lüdtke, O.; Robitzsch, A. A comparison of linking methods for longitudinal designs with the 2PL model under item parameter drift. *PsyArXiv* **2024**. [CrossRef]
40. Robitzsch, A. Bias and linking error in fixed item parameter calibration. *AppliedMath* **2024**, *4*, 1181–1191. [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.