

## Article

# Estimating Local Structural Equation Models

Alexander Robitzsch <sup>1,2</sup> 

<sup>1</sup> IPN–Leibniz Institute for Science and Mathematics Education, Olshausenstraße 62, 24118 Kiel, Germany; robitzsch@ipn.uni-kiel.de

<sup>2</sup> Centre for International Student Assessment (ZIB), Olshausenstraße 62, 24118 Kiel, Germany

**Abstract:** Local structural equation models (LSEM) are structural equation models that study model parameters as a function of a moderator. This article reviews and extends LSEM estimation methods and discusses the implementation in the R package *sirt*. In previous studies, LSEM was fitted as a sequence of models separately evaluated as each value of the moderator variables. In this article, a joint estimation approach is proposed that is a simultaneous estimation method across all moderator values and also allows some model parameters to be invariant with respect to the moderator. Moreover, sufficient details on the main estimation functions in the R package *sirt* are provided. The practical implementation of LSEM is demonstrated using illustrative datasets and an empirical example. Moreover, two simulation studies investigate the statistical properties of parameter estimation and significance testing in LSEM.

**Keywords:** local structural equation modeling; confirmatory factor analysis; differentiation; dedifferentiation; invariance

## 1. Introduction

A structural equation model (SEM) is a statistical approach for analyzing multivariate data (Bartholomew et al. 2011; Bollen 1989; Browne and Arminger 1995; Jöreskog et al. 2016; Shapiro 2012; Yuan and Bentler 2007). These models relate a multivariate vector  $\mathbf{X} = (X_1, \dots, X_I)$  of observed  $I$  variables (also referred to as items or indicators) to a vector of latent variables (i.e., factors)  $\boldsymbol{\eta}$  of a dimension smaller than  $I$ . SEMs constrain the mean vector  $\boldsymbol{\mu}$  and the covariance matrix  $\boldsymbol{\Sigma}$  of the random variable  $\mathbf{X}$  as a function of an unknown parameter vector  $\boldsymbol{\theta}$ . By doing so, the mean vector is constrained as  $\boldsymbol{\mu}(\boldsymbol{\theta})$ , and the covariance matrix is constrained as  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ .

Local structural equation models (LSEM) study SEMs as a function of a univariate moderator variable Hildebrandt et al. (2009, 2016). The moderator variable is the age or time variable in most applications. LSEM has been mentioned as a general tool for assessing measurement invariance across age or other continuous indicators in social sciences (Dong and Dumas 2020; Han et al. 2019; Leitgöb et al. 2023). Note that LSEM has also been abbreviated as LOSEM Briley et al. (2015a, 2015b).

The LSEM method is particularly suited for studying differentiation or dedifferentiation hypotheses (see Hildebrandt et al. 2009 or Molenaar et al. 2010b). Differentiation hypotheses of intelligence and general scholastic abilities describe changes in the relationship between different cognitive abilities (i.e., their structural organization) depending on the level of general ability (ability differentiation), age (differentiation in children and adolescents; dedifferentiation in older adults), and their interaction. Breit et al. (2022) presented a systematic review of 33 reports with data from 51 studies with over 260,000 participants that examined differentiation effects. The findings indicated practically significant ability differentiation in children and adults, and significant age dedifferentiation in older adults, with effect sizes that implicate a practical significance of the effects. However, Breit et al. (2022) also showed that age differentiation in children and adolescents was not supported. Instead, small but negligible effect sizes were found for age dedifferentiation in adolescents.



**Citation:** Robitzsch, Alexander. 2023. Estimating Local Structural Equation Models. *Journal of Intelligence* 11: 175. <https://doi.org/10.3390/jintelligence11090175>

Received: 4 June 2023

Revised: 16 August 2023

Accepted: 22 August 2023

Published: 1 September 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/>).

The LSEM method has been extended to two moderator variables by [Hartung et al. \(2018\)](#). [Molenaar \(2021\)](#) proposed a semiparametric moderated factor modeling approach in which no assumption concerning the functional form between the moderator and the model parameters are imposed. In contrast to the original definition of LSEM ([Hildebrandt et al. 2009](#)), some model parameters are allowed to be invariant across the continuous moderator variable.

LSEM is closely related to moderated nonlinear factor analysis (MNFA; [Bauer 2017](#); [Curran et al. 2014](#); [Molenaar and Dolan 2012](#)). In MNFA, a functional form of SEM model parameters as a function of a single moderator (or multiple moderators) is imposed. In this sense, MNFA is often more confirmatory than LSEM. Nevertheless, differentiation hypotheses were also investigated by means of MNFA ([Molenaar et al. \(2010a, 2010b, 2011, 2017\)\)](#). A tutorial on how to apply MNFA using the R package OpenMx ([Boker et al. 2011](#)) was given by [Kolbe et al. \(2022\)](#). LSEM also bears a similarity to the approach of individual parameter change ([Oberski 2013](#); [Arnold et al. \(2020, 2021\)](#)). Variation in SEM model parameters can also be tested with score-based invariance tests ([Huth et al. 2022](#); [Merkle and Zeileis 2013](#); [Wang et al. 2014](#)).

LSEM has been implemented in the R package *sirt* ([Robitzsch 2023b](#)) as a wrapper to the popular SEM package *lavaan* ([Rosseel 2012](#)). Moreover, the R package *umx* ([Bates et al. 2019](#)) can also be utilized for LSEM estimation.

This article reviews and extends LSEM estimation methods and discusses the implementation in the R package *sirt*. In previous literature, LSEM was fitted as a sequence of models that are separately evaluated as each value of the moderator variables. In this article, a joint estimation approach is proposed that is a simultaneous estimation method across all moderator values and also allows some model parameters to be invariant with respect to the moderator. Sufficient detail on the core estimation functions in the *sirt* package is provided. The article also evaluates two significance testing approaches to assess whether the moderator values are related to a model parameter in two simulation studies. Finally, an empirical example demonstrates the usefulness of the LSEM methodology.

The remainder of this article is structured as follows. Section 2 overviews the most important LSEM applications in the literature. In Section 3, different LSEM estimation and significance testing approaches are presented. Details about LSEM implementation in the *sirt* package can be found in Section 4. Section 5 discusses R input code and R output of an LSEM analysis involving illustrative datasets. Section 6 includes a simulation study investigating parameter recovery in LSEM regarding bias and root mean square error. Section 7 includes a simulation study that investigates different estimators of variability in parameter curves and the statistical properties of significance tests of parameter variation. In Section 8, an empirical example is presented that reanalyzes SON-R intelligence data for children aged between 21/2 and 7 years. Finally, Section 9 closes with a discussion.

## 2. Review of LSEM Applications

We now review important LSEM applications to demonstrate that this method is widely applied in substantive research. The original LSEM publication of [Hildebrandt et al. \(2009\)](#) (“Complementary and competing factor analytic approaches for the investigation of measurement invariance”) has been cited 93 times and 80 times, according to Google Scholar and ResearchGate (accessed on 18 July 2023), respectively. The second methodological LSEM publication by [Hildebrandt et al. \(2016\)](#) (“Exploring factor model parameters across continuous variables with local structural equation models”) has been cited 111 times, 89 times, and 77 times, according to Google Scholar, ResearchGate, and Web of Science (accessed on 18 July 2023), respectively. Hence, one could say that LSEM fills some niche in the researcher’s methodological toolbox.

In the following, some LSEM applications are briefly described. The studies are loosely organized according to the fields of application.

[Olaru and Allemand \(2022\)](#) examined differential and correlated change in personality across the adult lifespan using LSEM. [Brandt et al. \(2022\)](#) applied LSEM to four waves

of data obtained with the full NEO Personality Inventory collected over 11 years from 1667 adults in a US sample using age as a continuous moderator. [Hartung et al. \(2021\)](#) investigated the age-moderated covariance structure of the satisfaction with life scale (SWLS) and the domains of health satisfaction and financial satisfaction using LSEM. [Olaru et al. \(2019\)](#) analyzed NEO personality indicators across ages between 16 and 66 years by means of LSEM. They selected items for short scales that had the greatest extent of measurement invariance across age. [Seifert et al. \(2022\)](#) studied whether the rank-order stability of personality increases until midlife and declines later in old age and found that this inverted U-shaped pattern was not consistently observed in two reanalyses utilizing LSEM. Loneliness across different age levels was investigated by LSEM in [Entringer and Gosling \(2022\)](#) and [Panayiotou et al. \(2022\)](#). [Van den Akker et al. \(2021\)](#) applied LSEM for students aged between 8 and 18 years to investigate whether levels of conscientiousness and agreeableness decrease when levels of neuroticism increase, indicating a dip in personality maturation. [Gnambs \(2013\)](#) applied LSEM in a multitrait multi-informant meta-analysis for the big five factors.

[Hartung et al. \(2022\)](#) investigated the structure of the “dark personality factor” across age and gender using LSEM. [Krasko and Kaiser \(2023\)](#) investigated measurement invariance across age for the dark triad by means of LSEM.

[Bratt et al. \(2018\)](#) investigated levels of perceived age discrimination across early to late adulthood by employing LSEM, using data from the European social survey (ESS) collected in 29 countries. [Dutton and Kirkegaard \(2022\)](#) applied LSEM to investigate a particular question about the association between religiousness and intelligence. [Allemand et al. \(2022\)](#) used LSEM to investigate the effects of continuous age and COVID-19 virus worry on mean levels and correlations between gratitude and remaining opportunities and time. [Allemand et al. \(2021\)](#) examined age-related psychometrics and differences in the measurement, mean-levels, variances, and correlations of gratitude and future time perspective across adulthood using data in a representative Swiss sample for participants aged between 19 and 98 years.

[Schroeders et al. \(2015\)](#) studied the differentiation fluid and crystallized intelligence in German students of grades 5 to 12. [Watrin et al. \(2022\)](#) studied the age differentiation hypothesis of declarative knowledge, as proposed in Cattell’s investment theory. [Hülür et al. \(2011\)](#) studied with LSEM whether cognitive abilities become more differentiated with increasing age during childhood for children from age 2.5 to 7. [Hartung et al. \(2020\)](#) tested whether associations among executive functions strengthened from middle childhood to adolescence using cross-sectional data from a sample of children aged between 7 and 15 years. [Gnambs and Schroeders \(2020\)](#) examined the effects of cognitive abilities on the factor structure of the Rosenberg self-esteem scale across age by means of LSEM. [Whitley et al. \(2016\)](#) explored cross-sectional associations of age with five cognitive tests (word recall, verbal fluency, subtraction, number sequence, and numerical problem solving) in a large representative sample aged between 16 and 100 living in the UK. [Breit et al. \(2020\)](#) investigated ability differentiation, developmental differentiation, and their interaction with LSEM in two studies. [Breit et al. \(2021\)](#) provided a review of the literature on ability and developmental differentiation effects in children and youths. [Breit et al. \(2023\)](#) studied ability differentiation, including creativity measures, through LSEM for German students aged between 12 and 16 years.

[Hildebrandt et al. \(2010\)](#) employed LSEM to investigate structural invariance and age-related performance differences in face cognition. [Hildebrandt et al. \(2013\)](#) studied the specificity of face cognition compared with object cognition from individual differences and aging perspective by determining the amount of overlap between these abilities at the level of latent constructs across age. By utilizing LSEM, [Liu et al. \(2022\)](#) found that individual differences in white matter microstructure of the face processing brain network were more differentiated from global fibers with increasing ability.

LSEM was also applied in behavioral neurosciences [Kaltwasser et al. \(2017\)](#). [Jokić-Begić et al. \(2019\)](#) used LSEM for assessing measurement invariance across age for cyper-

chondria, a process of increased anxiety over one's health as a result of excessive online searching. Lodi-Smith et al. (2021) found that autism characteristics measured by the autism-spectrum quotient scale were not strongly associated with age by utilizing LSEM. Cox et al. (2016) used LSEM to quantify microstructural properties of the human brain's connections for understanding normal ageing and disease (see also Briley et al. (2015b)). Researchers de Mooij et al. (2018) used LSEM to study differences within and between brain and cognition across the adult life span. Zheng et al. (2019) investigated whether genetic and environmental influences on achievement goal orientations shift were moderated with age. Madole et al. (2019) applied LSEM in network analysis as a method for investigating symptom-level associations that underlie comorbidity connecting diagnostic syndromes.

Olaru et al. (2019) utilized LSEM in combination with ant colony optimization (see also Olaru and Jankowsky 2022) to resample and weight subjects to study differences in the measurement model across age as a continuous moderator variable.

An overview of different modeling strategies of LSEM for longitudinal data is presented in Olaru et al. (2020). Wagner et al. (2019) investigated through LSEM whether personality becomes more stable with age. They disentangled state and trait effects for the big five across the life span by applying LSEM to trait-state-occasion models. Gana et al. (2023) applied trait-state-occasion models in tandem with LSEM to investigate whether the characteristics of the depression EURO-D scale were associated with age.

LSEM was also applied to moderator variables different from age. Klieme and Schmidt-Borcherding (2023) employed LSEM to explore whether there is noninvariance for indicators of research self-efficacy regarding different training levels of students operationalized as the number of studied semesters. Weiss et al. (2020) investigated the threshold hypothesis of creativity by handling intelligence as a continuous moderator in LSEM. Schroeders and Jansen (2022) studied by means of LSEM whether the multidimensional structure of the science self-concept is moderated by levels of the cognitive ability in science. Basarkod et al. (2023) investigated whether reading self-concept dimensions vary across reading achievement levels in the PISA study. Olaru et al. (2022) examined the effects of family background on children's receptive vocabulary using LSEM with latent growth curve models. Bolsinova and Molenaar (2019) (see also Bolsinova and Molenaar 2018) used LSEM for indicator-specific covariates and extended LSEM to the study of cognitive tests involving reaction times.

### 3. Estimating and Testing Local Structural Equation Models

#### 3.1. Single-Group Structural Equation Model

In SEM, a measurement model is imposed that relates the observed variables  $X$  to latent variables  $\eta$

$$X = \nu + \Lambda\eta + \epsilon. \quad (1)$$

In addition, the covariance matrix of  $\epsilon$  is denoted by  $V$ ; that is,  $\text{Var}(\epsilon) = \Psi$ . Moreover,  $\eta$  and  $\epsilon$  are multivariate normally distributed random variables. In addition,  $\eta$  and  $\epsilon$  are assumed to be uncorrelated. In CFA, the multivariate normal (MVN) distribution is represented as  $\eta \sim \text{MVN}(\alpha, \Phi)$  and  $\epsilon \sim \text{MVN}(0, \Psi)$ . As we are only concerned with the covariance structure in SEM in this paper, we assume  $\alpha = 0$  and  $E(X) = \nu$ . Then, the covariance matrix of  $X$  in CFA can be computed as:

$$\text{Var}(X) = \Sigma(\theta) = \Lambda\Phi\Lambda^\top + \Psi. \quad (2)$$

The parameter vector  $\theta$  contains parameters in  $\Lambda$ ,  $\Phi$ , and  $\Psi$  that are estimated. Typically, the covariance matrix  $\Sigma$  is a constrained matrix determined by the specification (2).

In a general SEM, relationships among the latent variables  $\eta$  are modeled in path models. A matrix  $B$  of regression coefficients is specified such that:

$$\eta = B\eta + \zeta, \quad (3)$$



where  $\eta$  denotes an endogeneous and  $\zeta$  an exogeneous multivariate normally distributed latent variables. Note that (3) can be written as:

$$\eta = (I - B)^{-1} \zeta, \quad (4)$$

where  $I$  denotes the identity matrix. In this case, the covariance matrix of  $X$  are represented in SEM as:

$$\text{Var}(X) = \Sigma(\theta) = \Lambda(I - B)^{-1} \Phi[(I - B)^{-1}]^T \Lambda^T + \Psi. \quad (5)$$

Some identification constraints must be imposed when estimating the covariance structure of the SEM in (2) or (5) (Bollen 1989; Bollen and Davis 2009). The purpose of identifying constraints primarily lies in a convenient interpretation of latent variables  $\eta$  and is not primarily driven by improving the efficiency of estimating  $\Sigma$ .

When modeling multivariate normally distributed data without missing data, the empirical covariance matrix  $S$  is a sufficient statistic for the unknown covariance matrix  $\Sigma$ . Hence,  $S$  is also sufficient for the parameter vector  $\theta$  of the SEM in (2) or (5).

### 3.2. Multiple-Group Structural Equation Model

We now describe the general estimation of a multiple-group SEM. There exist  $G$  known groups  $g = 1, \dots, G$ . The allocation of a group to a subject is known in this case. Assume that group  $g$  has  $N_g$  subjects and an empirical covariance matrix  $S_g$ . The population covariance matrices are denoted by  $\Sigma_g$  ( $g = 1, \dots, G$ ). The model-implied covariance matrices are denoted by  $\Sigma_g(\theta)$  ( $g = 1, \dots, G$ ). The unknown parameter vector  $\theta$  can have common parameters across groups and parameters that are group-specific. For example, in a CFA, equal factor loadings and item intercepts across groups are frequently imposed (i.e., measurement invariance holds; Meredith 1993; Putnick and Bornstein 2016) by assuming the same loading matrix  $\Lambda$  across groups, while covariance matrices of latent variables or the matrix  $B$  of regression coefficients are allowed to differ across groups.

Up to constants, the maximum likelihood (ML) fitting function of the unknown parameter  $\theta$  for the covariance structure in the multiple-group SEM is given by (see Bollen 1989 and Jöreskog et al. 2016):

$$F(\theta; \{S_g\}_g) = \sum_{g=1}^G N_g \left( \log|\Sigma_g(\theta)| + \text{tr}(S_g \Sigma_g(\theta)^{-1}) - \log|S_g| - I \right). \quad (6)$$

Note that  $I$  refers to the number of observed variables; that is, the dimension of  $X$ . The set  $\{S_g\}_g$  denotes the set of  $G$  empirical covariance matrices that are sufficient statistics in multiple-group SEM estimation. The parameter vector  $\theta$  is estimated by minimizing  $F$  in (6) and is denoted as the ML estimate. The estimated parameter is denoted by  $\hat{\theta}$ .

In practice, the model-implied covariance matrix can be misspecified (Boos and Stefanski 2013; Gouriéroux et al. 1984; Kolenikov 2011; White 1982), and  $\theta$  is a pseudo-true parameter defined as the minimizer of the fitting function  $F$  in (6). Importantly,  $\theta$  does not refer to a parameter of the data-generating model in this case. In contrast, it should be interpreted as a summary of the data that are of central interest to the researcher.

The ML fitting function (6) can be considered a special case of discrepancy function. To this end, we define a general discrepancy function  $\mathcal{D}(S, \Sigma)$  between an empirical covariance matrix  $S$  and a population covariance matrix  $\Sigma$ . The real-valued nonnegative function  $\mathcal{D}$  should only attain the value zero if  $S = \Sigma$  (i.e., for correctly specified models). For the ML fitting function, the discrepancy function  $\mathcal{D}$  is defined as:

$$\mathcal{D}(S, \Sigma(\theta)) = \log|\Sigma(\theta)| + \text{tr}(S \Sigma(\theta)^{-1}) - \log|S| - I. \quad (7)$$

Using definition (7), we can rewrite (6) as:

$$F(\theta; \{S_g\}_g) = \sum_{g=1}^G N_g \mathcal{D}(S_g, \Sigma_g(\theta)) , \quad (8)$$

and  $\hat{\theta}$  is the minimizer of  $F(\theta; \{S_g\}_g)$ .

If an age moderator variable  $A$  is available, an SEM can, in principle, be estimated for all subgroups of subjects for different values of the age variable. In practice, sample sizes for concrete age values might be too small for separate estimation of the SEM. Moreover, discretizing the values of a continuous moderator variable  $A$  into  $G$  distinct groups of subjects might not be preferred due to loss of information (Hildebrandt et al. 2009). To circumvent these issues, LSEM has been proposed. We discuss LSEM estimation methods in the next subsections.

### 3.3. Local Weighting

Instead of grouping subjects that fall within a given range of the moderator, as in multiple-group SEMs, observations are locally weighted around focal points (i.e., specific values of the continuous moderator variable) in LSEM. In previous studies, SEMs are sequentially estimated on the basis of weighted samples of observations at all focal points (i.e., the pointwise LSEM estimation approach, see Section 3.5).

In LSEM, researchers are interested in investigating moderator-specific covariance structures. That is, they aim to model conditional covariances:

$$\text{Var}(X|A = a) = \Sigma(a) \quad (9)$$

As argued in the previous section, sample sizes might be too small for estimating  $\Sigma(a)$  only for subjects with  $A = a$ . To this end, subjects with moderator values  $a$  sufficiently close to a focal point  $a_t$  (i.e., a chosen value of the moderator variable  $A$ ) should also enter the estimation. For each focal point  $a_t$  and each subject  $n$ , weights  $w_{nt}$  are computed that reflect the distance of the moderator value (e.g., a value of age) of person  $n$  (i.e.,  $a_n$ ) and the focal point  $a_t$ . If  $a_n = a_t$ , the weight should be one, and it should be zero for age values  $a_n$  that strongly differ from  $a_t$ .

The computation of weights relies on a kernel function  $K$  that is chosen by the researcher (Hildebrandt et al. (2009, 2016)). The real-valued kernel function fulfills the properties  $K(0) = 1$ ,  $K(x) = K(-x)$  (i.e., it is a symmetry function),  $K(x) \geq 0$  for all  $x \in \mathbb{R}$ , and  $K$  is a decreasing function for  $x \geq 0$ . The subject-specific weight  $w_{nt}$  for subject  $n$  at a focal point  $a_t$  with a pre-specified bandwidth  $bw$  is computed as:

$$w_{nt} = K\left(\frac{a_n - a_t}{bw}\right) . \quad (10)$$

By the definition of  $K$ , weights are bounded within the interval  $[0, 1]$ .

Typical choices of the weight function in the literature of nonparametric regression or density estimation are the Gaussian kernel, the Epanechnikov kernel, and the uniform kernel function. The Gaussian kernel function is defined as:

$$K(x) = \exp(-x^2/2) . \quad (11)$$

In density estimation involving the Gaussian kernel function, an optimal bandwidth is given by  $bw = hN^{-1/5}\sigma_A$  with  $h = 1.1$ , and  $\sigma_A$  is the standard deviation of the age moderator variable (Silverman 1986). The parameter  $h$  is referred to as the bandwidth factor in this article. The Epanechnikov kernel function is defined as:

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{for } |x| \leq 1 \\ 0 & \text{for } |x| > 1 \end{cases} . \quad (12)$$

For age values  $a_n$  with  $|a_n - a_t| > bw$ , weights  $w_{nt}$  are zero. Finally, the uniform kernel function is defined as:

$$K(x) = \begin{cases} 1 & \text{for } |x| \leq 1 \\ 0 & \text{for } |x| > 1 \end{cases} \quad (13)$$

The uniform kernel can be used to define weights so that they reflect the discretization of the continuous age variable  $A$  into  $G$  distinct groups. The estimated LSEM will provide parameter results that are identical to the multiple-group SEM if the same identification constraints are utilized.

### 3.4. Estimation of Conditional Means and Conditional Covariances

We now describe the estimation conditional covariances  $\Sigma(a)$ . In a practical implementation of the LSEM, researchers define a discrete grid of moderator values  $a_1, a_2, \dots, a_T$  (i.e., the focal points) of the age variable  $A$ . In most applications, a grid of equidistant focal points is chosen (Hildebrandt et al. 2016). However, the grid of focal points could also be chosen in such a way that it mimics the empirical distribution of the moderator variable. For example, researchers might use empirical percentiles of the moderator variable (e.g., a grid of 10 focal points using the  $p$ th percentile for  $p = 5, 15, \dots, 95$ ).

To estimate conditional covariances at a focal point  $a_t$ , we first compute the conditional mean function  $E(X|A = a)$  for  $X = (X_1, \dots, X_I)$ . For a variable  $X_i$  for  $i = 1, \dots, I$ , a local quadratic regression model is specified to estimate the conditional mean at focal point  $a_t$ . That is, one minimizes:

$$(\hat{\gamma}_{it0}, \hat{\gamma}_{it1}, \hat{\gamma}_{it2}) = \arg \min_{(\gamma_{it0}, \gamma_{it1}, \gamma_{it2})} \left\{ \sum_{n=1}^N w_{nt} \left( x_{in} - \gamma_{it0} - \gamma_{it1}(a_{nt} - a_t) - \gamma_{it2}(a_{nt} - a_t)^2 \right)^2 \right\} \quad (14)$$

The conditional mean estimate of  $\mu_i(a_t) = E(X_i|A = a_t)$  is given by  $\hat{\mu}_i(a_t) = \hat{\gamma}_{it0}$ . Note that the minimization in (14) is a weighted least squares estimation problem for a linear regression (i.e., it is linear in model parameters) and closed formulae are available for estimating  $(\gamma_{it0}, \gamma_{it1}, \gamma_{it2})$  (see Fox 2016).

We now describe the estimation of conditional covariances  $\sigma_{ij}(a_t) = \text{Cov}(X_i, X_j|A = a_t)$ . First, residuals  $e_{nit}$  are computed using local quadratic regression parameters defined in (14) as:

$$e_{nit} = x_{in} - \hat{\gamma}_{it0} - \hat{\gamma}_{it1}(a_{nt} - a_t) - \hat{\gamma}_{it2}(a_{nt} - a_t)^2. \quad (15)$$

The estimate of the conditional covariances  $\sigma_{ij}(a_t)$  can be obtained by simple weighting or a local regression model.

In the weighting approach, one estimates:

$$\hat{\sigma}_{ij}(a_t) = W_t^{-1} \sum_{n=1}^N w_{nt} e_{nit} e_{njt}, \quad (16)$$

where  $W_t = \sum_{n=1}^N w_{nt}$ . This approach was advocated in Hildebrandt et al. (2009) and Hildebrandt et al. (2016).

In recently proposed local regression modeling (see Olaru et al. 2020), one also specifies a local quadratic regression estimation problem for the computation of the conditional covariance:

$$(\hat{\delta}_{ijt0}, \hat{\delta}_{ijt1}, \hat{\delta}_{ijt2}) = \arg \min_{(\delta_{ijt0}, \delta_{ijt1}, \delta_{ijt2})} \left\{ \sum_{n=1}^N w_{nt} \left( e_{nit} e_{njt} - \delta_{ijt0} - \delta_{ijt1}(a_{nt} - a_t) - \delta_{ijt2}(a_{nt} - a_t)^2 \right)^2 \right\}. \quad (17)$$

The estimate of the conditional covariance is given as  $\hat{\sigma}_{ij}(a_t) = \hat{\delta}_{ijt0}$ .

Note that the estimation of the conditional mean function in (14) and the conditional covariance function in (17) is essentially equivalent, except for the case that the former uses the values  $x_{ni}$  as the dependent variable  $x_{ni}$  (i.e., indicator  $i$ ), while the latter uses

the product residual  $e_{nit}e_{njt}$  of variables for indicators  $i$  and  $j$  for the computation of the moderator-specific conditional covariance.

The steps can be repeated for all pairs of variables  $i$  and  $j$  ( $i, j = 1, \dots, I$ ) and all focal points  $a_t$  ( $t = 1, \dots, T$ ). The resulting estimated conditional covariance matrices at focal points  $a_t$  are denoted by  $\hat{\Sigma}_t$  ( $t = 1, \dots, T$ ). The estimated covariance matrices  $\hat{\Sigma}_t$  are not guaranteed to be positive definite. Therefore, the estimate might be slightly modified to determine a close matrix to  $\hat{\Sigma}_t$  that fulfills the positive definiteness property (Bentler and Yuan 2011).

LSEM estimation methods rely on the estimated conditional covariances. Three different estimation approaches are described in Sections 3.5, 3.6 and 3.8.

### 3.5. Pointwise LSEM Estimation

Pointwise LSEM estimation relies on the idea that a separate SEM is fitted to each focal point  $a_t$ . The resulting parameter estimates  $\hat{\theta}_t$  are plotted or analyzed as a function of the age variable  $A$ . More formally, based on the conditional covariance estimate  $\hat{\Sigma}_t$ , at each focal point  $a_t$ , the following fitting function is minimized:

$$F(\theta_t; \hat{\Sigma}_t) = \mathcal{D}(\hat{\Sigma}_t, \Sigma_t(\theta_t)) , \quad (18)$$

where  $\hat{\theta}_t$  denotes the minimizer of  $F(\theta_t; \hat{\Sigma}_t)$ . Note that in (18), the distance between the empirical conditional covariance  $\hat{\Sigma}_t$  and the model-implied conditional covariance  $\Sigma_t(\theta_t)$  at the focal point  $a_t$  is minimized. This approach was proposed by Hildebrandt et al. (2009, 2016). The minimization in (18) is not restricted to ML estimation and can also be applied to weighted least estimation in SEM (Browne 1974) or model-robust fitting functions (Robitzsch 2023a).

Model fit statistics, such as RMSEA, SRMR, or TLI, are computed at each value of the focal point. Note that pointwise LSEM estimation provides parameter curves across different values of the moderator variable.

The pointwise LSEM estimation method allows the parameter vector  $\theta(a)$  to vary freely across  $a$ . However, this flexibility sometimes hinders interpretation. Moreover, some researchers might prefer to impose invariance constraints for some of the model parameters (Leitgöb et al. 2023). For this reason, a joint LSEM estimation approach is proposed that is described in the next Section 3.6.

### 3.6. Joint LSEM Estimation with Invariance Constraints

While pointwise LSEM estimation tackles the estimation problem by successively and separately estimating an SEM at each of the focal points, joint LSEM estimation defines a single estimation function that involves conditional covariance matrices of all focal points. By doing so, the parameter vector  $\theta$  can contain parameters that are specific to each focal point and parameters that do not vary for different values of age. The fitting function is defined as:

$$F(\theta; \{\hat{\Sigma}_t\}_t) = \sum_{t=1}^T W_t \mathcal{D}(\hat{\Sigma}_t, \Sigma_t(\theta)) , \quad (19)$$

where  $\hat{\theta}$  is the minimizer of  $F(\theta; \{\hat{\Sigma}_t\}_t)$  and  $W_t = \sum_{n=1}^N w_{nt}$  is the sum of weights specific to each focal point  $a_t$ . Note that (19) looks like a fitting function in multiple-group SEM estimation. However, subjects can enter multiple groups (i.e., focal points) because they enter the estimated conditional covariances multiple times according to the weights  $w_{nt}$ . Hence, the fitting function  $F$  in (19) will not be an ML fitting function and falls in the general class of M-estimation problems (Stefanski and Boos 2002).



The parameter vector  $\theta$  can be decomposed into components  $\theta = (\theta_0, \theta_1, \dots, \theta_T)$ , where  $\theta_0$  contains parameters that are invariant across age, and  $\theta_t$  for  $t \geq 1$  contain the parameters that vary across age values. The fitting function in (19) can then be rewritten as:

$$F(\theta_0, \theta_1, \dots, \theta_T; \{\hat{\Sigma}_t\}_t) = \sum_{t=1}^T W_t \mathcal{D}(\hat{\Sigma}_t, \Sigma_t(\theta_0, \theta_t)) . \quad (20)$$

Note that the originally proposed pointwise estimation of the fitting function in (18) is equivalent to joint LSEM estimation in (20) if there does not exist invariant model parameters  $\theta_0$ .

In joint LSEM estimation, global model fit statistics are computed. These fit statistics can be interpreted similarly as in multiple-group SEMs.

### 3.7. Estimation of DIF Effects

In joint LSEM estimation defined by the fitting function  $F$  in (20), some parameters (i.e., the parameter vector  $\theta_0$ ) have invariance constraints across the age moderator variable. These invariance constraints ease interpretation and have the advantage of specifying parsimonious SEMs. However, researchers might be interested in what would happen if these invariance constraints were freed.

Violations of measurement invariance are referred to as differential item functioning (DIF) in item response theory literature (Mellenbergh 1989; Holland and Wainer 1993; Millsap 2011). Noninvariant parameters are referred to as DIF effects in this literature. We also use this notation and now discuss the estimation of DIF effects. DIF effects emerge if all estimated age-specific parameters  $\hat{\theta}_t$  ( $t \geq 1$ ) are held fixed in (20), and the entries of the parameter vector  $\theta_0$  are allowed to vary across age. We denote the focal-point-specific estimates of DIF effects by  $\delta_t$ . To this end, invariant parameters  $\theta_0$  are replaced with  $\delta_1, \dots, \delta_T$ , and the following fitting function  $F$  is minimized to obtain DIF effect estimates  $\hat{\delta}_t$  ( $t = 1, \dots, T$ ):

$$F(\delta_1, \dots, \delta_T, \hat{\theta}_1, \dots, \hat{\theta}_T; \{\hat{\Sigma}_t\}_t) = \sum_{t=1}^T W_t \mathcal{D}(\hat{\Sigma}_t, \Sigma_t(\delta_t, \hat{\theta}_t)) . \quad (21)$$

Note that there are no invariant model parameters in (21), and the DIF effects  $\delta_t$  at the focal point  $a_t$  could alternatively be obtained by pointwise minimization of:

$$F(\delta_t, \hat{\theta}_t; \hat{\Sigma}_t) = \mathcal{D}(\hat{\Sigma}_t, \Sigma_t(\delta_t, \hat{\theta}_t)) . \quad (22)$$

The estimated DIF effects can be plotted or analyzed as a function of the age moderator to investigate whether the invariance constraints are substantially violated.

### 3.8. Joint LSEM Estimation with More General Parameter Constraints and Relation to Moderated Nonlinear Factor Analysis

In this subsection, joint LSEM estimation is slightly generalized. The fitting function is the same as in (20), but constraints across focal-point-specific parameters  $\theta_t$  are allowed. In particular, we discuss the implementation of linear, quadratic, and piecewise linear or quadratic parameter constraints.

Assume a parameter curve  $\theta(a_t)$  for a particular parameter. Furthermore, assume that the focal points are equidistant; that is,  $a_{t+1} - a_t = \Delta$  are equal for  $t = 1, \dots, T - 1$ .

We first describe a linear parameter constraint. A linear function of a parameter  $\theta$  for age values  $a$  is given by  $f(a) = \alpha_0 + \alpha_1 a$ . The first derivative of  $f$  is constant, and it holds that  $f'(a_{t+1}) = f'(a_t) = \alpha_1$ . Hence, the equality in derivatives can be translated into equalities in first-order differences in model parameters:

$$\theta(a_{t+2}) - \theta(a_{t+1}) = \theta(a_{t+1}) - \theta(a_t) . \quad (23)$$

These constraints can be added in multiple-group SEM in typical SEM software such as lavaan (Rosseel 2012).

A quadratic function of a parameter is given by  $f(a) = \alpha_0 + \alpha_1 a + \alpha_2 a^2$ . This function has constant second-order derivatives; that is,  $f''(a_{t+1}) = f''(a_t) = 2\alpha_2$ . Hence, second-order differences in parameter values are constant, which translates into:

$$\theta(a_{t+3}) - 2\theta(a_{t+2}) + \theta(a_{t+1}) = \theta(a_{t+2}) - 2\theta(a_{t+1}) + \theta(a_t). \quad (24)$$

Similarly, cubic parameter constraints can be implemented by recognizing that the third-order differences in parameter values are constant. A slightly more tedious constraint than (24) can be derived.

The linearity and quadratic constraints in (23) and (24) can also be applied if parameter curves are broken into segments. Hence, piecewise linear or quadratic functions can be applied.

Applying (piecewise) quadratic parameter functions in joint LSEM estimation can be interpreted as a kind of smoothing procedure to stabilize parameter estimation. Furthermore, the raw data are smoothed when computing the estimated conditional covariance matrices  $\hat{\Sigma}_t$ . Hence, researchers have two choices for how stabilizing parameter estimation in LSEM.

Notably, parameter constraints in joint LSEM estimation are estimates of MNFA in a particular case. If the age moderator values  $A$  has only values at the grid of equidistant focal points  $a_1, \dots, a_T$ , then using the uniform kernel with  $bw = (a_2 - a_1)/2$  is equivalent to MNFA with appropriate parameter constraints. Such an approach is described in Tucker-Drob (2009).

### 3.9. Parameter Curve Summaries and Significance Testing

Finally, we discuss the definition of summary statistics and the test of significant parameter variation across age. Let  $\theta(a_t)$  be a parameter curve of some model parameter estimated at focal points  $a_t$  ( $t = 1, \dots, T$ ). The parameter curve  $\theta(a)$  can be summarized by the mean and the standard deviation. Let  $f(a_t)$  be the discrete density of the age variable  $A$  at focal point  $a_t$  and assume that  $\sum_{t=1}^T f(a_t) = 1$ . The (weighted) average value of the parameter curve (i.e., the mean) is given as:

$$M_{\theta(a)} = \sum_{t=1}^T f(a_t) \theta(a_t). \quad (25)$$

In practice, an estimate of (25) is obtained by

$$\hat{M}_{\theta(a)} = \sum_{t=1}^T \hat{f}(a_t) \hat{\theta}(a_t). \quad (26)$$

The standard deviation of a parameter curve quantifies the variability of a parameter curve across age and is given by:

$$SD_{\theta(a)} = \sqrt{\sum_{t=1}^T f(a_t) \left( \theta(a_t) - M_{\theta(a)} \right)^2}. \quad (27)$$

An estimate of the standard deviation defined in (27) is given by:

$$\widehat{SD}_{\theta(a)} = \sqrt{\sum_{t=1}^T \hat{f}(a_t) \left( \hat{\theta}(a_t) - \hat{M}_{\theta(a)} \right)^2}. \quad (28)$$

The sample estimate  $\widehat{SD}_{\theta(a)}$  is always positive in finite samples if no invariance constraints are imposed. Hence, the naive standard deviation estimate in (28) will be positively biased.

The bootstrap resampling procedure (Efron and Tibshirani 1994) can be used to reduce the bias in an estimate of  $SD_{\theta(a)}$ . For LSEM, nonparametric bootstrap is implemented, which resamples subjects with replacement. The pointwise standard deviation of a parameter value across bootstrap samples can be used as a standard error estimate. A bias-corrected estimate of the standard deviation is obtained by:

$$\widehat{SD}_{\theta(a),bc} = \text{sqr}_+( \widehat{SD}_{\theta(a)}^2 - \mathcal{B}_{\theta(a)} ) , \quad (29)$$

where  $\text{sqr}_+(x) = \sqrt{\max(x, 0)}$  and  $\mathcal{B}_{\theta(a)}$  is the finite-sample bias of  $\widehat{SD}_{\theta(a)}^2$  that can be determined by bootstrap resampling (Efron and Tibshirani 1994). A t-statistic for significant variation in an estimated parameter curve can be computed as:

$$t = \widehat{SD}_{\theta(a),bc} / SE , \quad (30)$$

where  $SE$  is the standard deviation of  $\widehat{SD}_{\theta(a)}$  values defined in (28) across different bootstrap samples. Note that this test procedure relies on a normal distribution assumption for the test statistic  $t$ , although it is probably an incorrect null distribution.

An alternative test for parameter variation is based on a Wald test. A covariance matrix estimate  $V$  for the vector  $\xi = (\theta(a_1), \dots, \theta(a_T))$  can be obtained from bootstrap. It is assumed that  $\hat{\xi}$  is multivariate normally distributed. Let  $H$  be a  $(T-1) \times T$  matrix that implements equality constraints across the values of the parameter curve. The null hypothesis of no parameter variation is given by  $H\xi = \mathbf{0}$ . Consider the Wald test statistic:

$$\chi^2 = \hat{\xi}^\top H^\top (H^\top V H)^{-1} H \hat{\xi} \quad (31)$$

This statistic is chi-square distributed with  $T-1$  degrees of freedom.

In previous work, a permutation test has been proposed for testing parameter variation (Hartung et al. 2022; Hildebrandt et al. (2009, 2016)). A permutation test simultaneously assesses the effects on all parameters. In contrast, the test based on the standard deviation (30) and the Wald test (31) relies on a fitted model without modifying all other model parameters. Hence, we tend to favor the latter statistics over the permutation test.

#### 4. Implementation of Local Structural Equation Models in the Sirt Package

In this section, we discuss the implementation of LSEM in the R (R Core Team 2023) package *sirt* (Robitzsch 2023b). The CRAN version can be installed within R using `utils::install.packages('sirt')`, while the most recent GitHub version can be installed employing `devtools::install_github('alexanderrobitzsch/sirt')`. The four primary LSEM functions are `sirt::lsem.estimate()`, `sirt::lsem.bootstrap()`, `sirt::lsem.test()` and `sirt::lsem.permutationTest()`, which will be discussed below. The new CRAN release of *sirt* from August 2023 (*sirt* 3.13-228; <https://cran.r-project.org/web/packages/sirt/> accessed on 11 August 2023) includes the functionality described in this article.

LSEM estimation in *sirt* provides a wrapper to the SEM package *lavaan* (Rosseel 2012). The model specification follows the *lavaan* syntax, which eases the familiarity with R code for LSEM estimation because *lavaan* seems to be the most popular open-source SEM software.

In Listing 1, the main function `sirt::lsem.estimate()` is displayed. This function is the main LSEM estimation function. We now discuss the most important arguments in detail.

**Listing 1.** LSEM function `sirt::lsem.estimate()`.

```

1 sirt::lsem.estimate(data, moderator, moderator.grid, lavmodel, type="LSEM", h=1.1, bw=NULL,
2 residualize=TRUE, fit_measures=c("rmsea", "cfi", "tli", "gfi", "srmr"),
3 standardized=FALSE, standardized_type="std.all", lavaan_fct="sem",
4 sufficient_statistics=TRUE, pseudo_weights=0, sampling_weights=NULL,
5 loc_linear_smooth=TRUE, est_joint=FALSE,
6 par_invariant=NULL, par_linear=NULL, par_quadratic=NULL,
7 partable_joint=NULL, pw_linear=1, pw_quadratic=1, pd=TRUE, est_DIF=FALSE,
8 se=NULL, kernel="gaussian", eps=1E-8, verbose=TRUE, ... )

```

In data, a data frame must be provided by the user. The data frame should also include the moderator variable, whose variable name must be specified in `moderator`. The set of focal points can be defined as a vector `moderator.grid`. In `lavmodel`, lavaan syntax must be provided for estimating the LSEM. The default of the argument `type` is ‘LSEM’; that is, an LSEM is estimated. By choosing `type='MGM'`, a multiple-group model with a discretized moderator variable is estimated. The bandwidth in `sirt::lsem.estimate()` can be specified by `h` or `bw`. The arguments are related through the formula:

$$bw = hN^{-1/5}\hat{\sigma}_A, \quad (32)$$

where  $\hat{\sigma}_A$  denotes the estimated standard deviation of the moderator variable  $A$  (i.e., the argument `moderator`). The logical argument `residualize` indicates whether local regression smoothing of the mean structure should be applied before estimating conditional covariances. The argument `fit_measures` defines fit statistics available in lavaan that should be included in the LSEM output. The logical argument `standardized` defines whether standardized parameters should appear in the LSEM output. The type of standardization is specified in `standardized_type` whose conventions follow the lavaan package. In `lavaan_fct`, the lavaan function is specified that is used for LSEM estimation. The default `lavaan_fct="sem"` refers to `lavaan::sem()`. Other options are `"cfa"` (for `lavaan::cfa()`) and `"lavaan"` (for `lavaan::lavaan()`). The logical argument `sufficient_statistics` indicates whether sufficient statistics (i.e., conditional mean and conditional covariances) should be used in estimation. Without missing data, ML can always rely on sufficient statistics. However, in the presence of missing data, conditional covariance matrices are estimated based on pairwise deletion. However, if full information maximum likelihood was utilized, the mean structure cannot be properly residualized. Hence, researchers are advised either to believe in missing data mechanisms close to missing completely at random that justify the usage of pairwise deletion or to apply an appropriate multiple imputation procedure prior to LSEM analysis if there are missing values in the dataset.

Users can also input a vector of sampling weights in `sampling_weights`. The logical argument `loc_linear_smooth` defines whether local quadratic regression (see (17)) should be applied in the estimation of conditional covariances. If the default `loc_linear_smooth=TRUE` is changed into `loc_linear_smooth=FALSE`, the weighting formula (16) is utilized. The logical argument `est_joint` indicates whether joint LSEM estimation (i.e., the default; see Sections 3.6 or 3.8) or pairwise LSEM estimation (see Section 3.5) is applied. Invariant model parameters can be specified in the vector argument `par_invariant`. If there are some invariant parameters, joint LSEM estimation is automatically chosen (i.e., `est_joint=TRUE`). Linear or quadratic parameter constraints on model parameters (see Section 3.8) can be specified with `par_linear` and `par_quadratic`, respectively. The number of segments in piecewise linear or piecewise quadratic parameter constrained estimation can be specified with `pw_linear` or `pw_quadratic`. The default is that the constraints should be applied across all moderator values (i.e., there is only one segment of a piecewise linear or quadratic function). The argument `partable_joint` allows the input of a lavaan parameter table in joint estimation. This argument has the advantage that arbitrary parameter constraints can be specified by the user (e.g., additional equality constraints in piecewise quadratic functions). The logical argument `pd` indicates whether non-positive definite conditional covariance

matrices should be smoothed to ensure positive definiteness. The logical argument `est_DIF` defines whether DIF effects should be estimated (see Section 3.7). Note that DIF effects can only be estimated if the LSEM model contains some invariant model parameters. The argument `kernel` allows the choice of the kernel function. Possible options are “gaussian”, “epanechnikov”, and “uniform”. Finally, the logical argument `verbose` indicates whether some output should be displayed in the R console when estimating the LSEM model.

Listing 2 displays the LSEM bootstrapping function in the `sirt` package. An object must be provided that is the output of the `sirt::lsem.estimate()` function. The number of bootstrap samples can be specified by the argument `R`. Bootstrap can also be applied at the level of higher-order units. For example, school classes, schools, or organizations can be bootstrapped instead of bootstrapping subjects. Such a kind of cluster bootstrap is required if there is an additional dependency structure in the data. In this case, users can define a vector of cluster units in `cluster`. The `sirt::lsem.bootstrap()` also allows more general replication designs such as jackknife, balanced repeated replication, or half sampling (Kolenikov 2010) by providing an  $N \times R$  matrix of resampling weights in the argument `repl_design`. In the case of more complex designs, a scale factor `repl_factor` must be defined by the user for a correct standard error computation. In the case of jackknife, it is 1 (or  $(R - 1)/R$ ), while it is  $1/R$  in the case of bootstrap resampling. The bootstrap function `sirt::lsem.bootstrap()` is needed for computing the standard deviation statistic of parameter curves and its statistical inference (see Section 3.9). The `sirt::lsem.bootstrap()` function also allows an option for parallel computing. The number of employed cores can be specified by the argument `n.core`. The default is the use of one core which means that no parallel computing is applied in LSEM bootstrap estimation.

**Listing 2.** LSEM function `sirt::lsem.bootstrap()`.

```
1 sirt::lsem.bootstrap(object, R=100, verbose=TRUE, cluster=NULL,
2 repl_design=NULL, repl_factor=NULL, use_starting_values=TRUE,
3 n.core=1, cl.type="PSOCK")
```

Listing 3 displays the LSEM function `sirt::lsem.test()` that performs the Wald tests for parameter variation (see Section 3.9). Instead of applying a test of the equality of a parameter curve on  $T$  focal points  $a_1, \dots, a_T$ , the specification in `models` allows the test of significant regression parameters for a particular function. For example, a specification “`FX=~X1`” =  $y \sim m + I(m^2)$  tests whether the vector of the linear and the quadratic regression coefficient of the factor loading `FX=~X1` differs from  $(0,0)$ . Note that `sirt::lsem.test()` requires the output of `sirt::lsem.estimate()` in `mod` and the output of the application of the bootstrap (or general resampling) of `sirt::lsem.bootstrap()` in `bmod`.

**Listing 3.** LSEM function `sirt::lsem.test()`.

```
1 sirt::lsem.test(mod, bmod, models=NULL )
```

Listing 4 displays the LSEM function `sirt::lsem.permutationTest()` that carries out the permutation test for a statistical significance test for variation in parameter curves of the LSEM model Hildebrandt et al. (2009, 2016). In the permutation test, the values of the moderator variables are randomly resampled in the dataset to create a null distribution of parameter curves under the assumption of no relation to the moderator. The number of permutation samples can be specified in the argument `B`. As in `sirt::lsem.bootstrap()`, parallel computing can be requested by the number of cores in the argument `n.core`.



**Listing 4.** LSEM function `sirt::lsem.permutationTest()`.

```

1 sirt::lsem.permutationTest(lsem.object, B=1000, residualize=TRUE, verbose=TRUE,
2 n.core=1, cl.type="PSOCK")

```

## 5. Illustrative Datasets

In this section, we illustrate LSEM estimation with the R package *sirt*. Three simulated datasets involving six variables  $x_1$ ,  $x_2$ ,  $x_3$ ,  $y_1$ ,  $y_2$ , and  $y_3$  are used for illustration. The analysis model is a two-dimensional factor model with a simple loading structure, where the first factor  $FX$  is measured by  $x_1$ ,  $x_2$ , and  $x_3$ , and the second factor  $FY$  is measured by  $y_1$ ,  $y_2$ , and  $y_3$ . The moderator variable *age* was assessed at 13 time points, referring to ages 6, 7, ..., 18. An anonymous reviewer pointed out that using 13 time points would look like longitudinal data. However, we only used the 13 time points for illustratory purposes. For example, there could be 13 cross-sectional age groups that are assessed.

The population parameters of the factor model for each age  $a = 6, 7, \dots, 18$  and each of the three datasets *DATA1*, *DATA2*, and *DATA3* can be found in the directory “*POPPARS*” at [https://osf.io/puaz9/?view\\_only=63ffb2fd30f5400e89c59d03366bf793](https://osf.io/puaz9/?view_only=63ffb2fd30f5400e89c59d03366bf793) (accessed on 3 June 2023). From these population parameters, 10,000 subjects were simulated at each of the 13 age points. The distribution at each age point exactly coincides with the specified conditional mean vector and the conditional covariance matrix (see, e.g., the `lavaan::simulateData()` function with the argument `empirical=FALSE` for a similar functionality). Data were simulated from a multivariate normal distribution. This simulation ensures that the population data involving 130,000 subjects (i.e.,  $= 13 \times 10,000$  subjects) exactly follows the specified covariance structure. In *DATA1*, all model parameters except for residual variances were assumed noninvariant. In *DATA2*, only the structural parameters (i.e., factor correlation and factor variances) were noninvariant, while factor loadings and residual variances were assumed invariant. In *DATA3*, all measurement and structural model parameters were assumed invariant. The population datasets and the data-generating model parameters can be found in the directory “*POPDATA*” at [https://osf.io/puaz9/?view\\_only=63ffb2fd30f5400e89c59d03366bf793](https://osf.io/puaz9/?view_only=63ffb2fd30f5400e89c59d03366bf793) (accessed on 3 June 2023). The illustrative datasets used in this section were subsamples of 2000 subjects from datasets *DATA1*, *DATA2*, and *DATA3*. The main motivation for using a subsample of the data is to show that LSEM produces some variability in model parameter estimates even if the model parameter is invariant across the moderator values in the data-generating model. The subsamples were created by random sampling without replacement from the population datasets. These datasets can be found in the directory “*ILLUSDATA*” at [https://osf.io/puaz9/?view\\_only=63ffb2fd30f5400e89c59d03366bf793](https://osf.io/puaz9/?view_only=63ffb2fd30f5400e89c59d03366bf793) (accessed on 3 June 2023).

Listing 5 contains the specification of the LSEM model involving two factors  $FX$  and  $FY$ . In lines 5–10 in Listing 5, the *lavaan* syntax for the factor model is specified in the string `lavmodel1`. Line 13 in Listing 5 defines the parameter names (i.e., the factor loadings of  $x_2$ ,  $x_3$ ,  $y_2$ , and  $y_3$ ) that are assumed invariant across the values of the moderator variable *age*. Line 16 in Listing 5 specifies the vector of focal points at which the LSEM model should be estimated. Lines 19–21 in Listing 5 contain the R command for applying `sirt::lsem.estimate()`. Note that the invariant model parameters are provided with the argument `par_invariant`, DIF effects were estimated due to `est_DIF=TRUE`, and the bandwidth factor  $h$  was chosen as 1.1. Joint LSEM estimation was applied because invariance constraints among parameters were imposed. In line 25 in Listing 5, the random seed is fixed, which ensures that bootstrap resampling will not change when applying code at a different time. Line 26 in Listing 5 specifies bootstrapping using `sirt::lsem.bootstrap()`. In total,  $R = 200$  bootstrap samples were utilized. Note that the specified factor model in Listing 5 is misspecified for the dataset *DATA1*, but correctly specified for the datasets *DATA2* and *DATA3*.

**Listing 5.** Illustrative datasets: Specification of LSEM with invariant factor loadings in `sirt::lsem.estimate()` and subsequent bootstrap in `sirt::lsem.bootstrap()`.

```

1 library(lavaan)
2 library(sirt)
3
4 # specify model using lavaan syntax
5 lavmodel <- "
6 FX=~ 1*X1+X2+X3
7 FY=~ 1*Y1+Y2+Y3
8 FX ~~ FX
9 FY ~~ FY
10 FX ~~ FY"
11
12 #- define invariant parameters
13 par_invariant <- c("FX=~X2", "FX=~X3", "FY=~Y2", "FY=~Y3" )
14
15 #- define grid of moderator values
16 moderator.grid <- seq(6,18,1)
17
18 # estimate LSEM model
19 mod <- sirt::lsem.estimate(dat, moderator="age", moderator.grid=moderator.grid,
20 sufficient_statistics=TRUE, lavmodel=lavmodel, h=1.1,
21 par_invariant=par_invariant, standardized=TRUE, est_DIF=TRUE)
22 summary(mod) # print summary
23
24 # perform bootstrap with R=200 bootstrap samples
25 set.seed(789)
26 rmod <- sirt::lsem.bootstrap(mod, R=200)
27 summary(rmod)

```

A part of R output of the `sirt::lsem.bootstrap()` function can be found in Listing 5. A slight misfit is detected in fit statistics RMSEA and SRMR. The CFI and TLI fit statistics are not indicative of the incorrect invariance assumption of factor loadings.

Figure 1 displays parameter curves for the two factor variances (i.e.,  $FX \sim FX$  and  $FY \sim FY$ ) and the factor correlation (i.e.,  $std\_FX \sim FY$ ) for the illustrative dataset *DATA1*. From Listing 5, we see that the variance of *FX* had an average of 0.396 with significant parameter variation ( $SD_{bc} = 0.083$ ,  $p < 0.001$ ), and *FY* had an average of 0.473 with significant parameter variation ( $SD_{bc} = 0.111$ ,  $p < 0.001$ ). Moreover, the factor correlation had an average of 0.584 and also showed a significant parameter variation ( $SD_{bc} = 0.059$ ,  $p = 0.003$ ).

Figure 2 displays parameter curves for the two factor variances and the factor correlation for the illustrative dataset *DATA3*, which had no simulated parameter variation in these parameters. By comparing Figures 1 and 2, it is evident that there is negligible parameter variation for the dataset *DATA3* compared to the dataset *DATA1*.

The parameter curves for DIF effects for factor loadings for datasets *DATA1* and *DATA2* are displayed in Figures 3 and 4, respectively. For *DATA1*, factor loadings were simulated as noninvariant, while they were assumed invariant across age for *DATA2*. This fact is visible when comparing Figures 3 and 4.

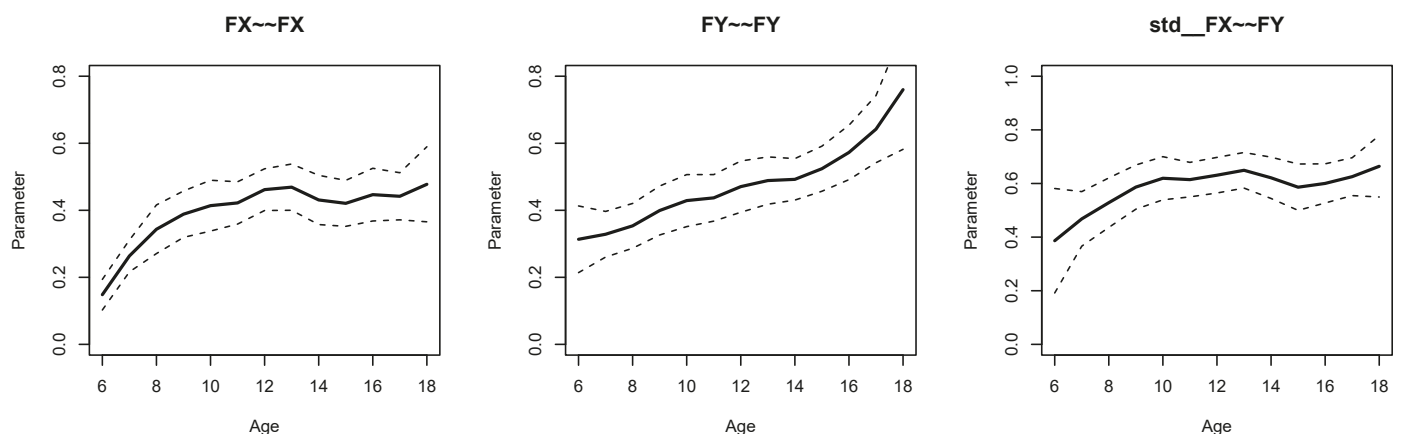
It can be seen from Listing 6 that DIF effects for factor loadings  $x_1$  ( $SD_{bc} = 0.024$ ,  $p = 0.020$ ),  $x_3$  ( $SD_{bc} = 0.038$ ,  $p = 0.002$ ),  $y_1$  ( $SD_{bc} = 0.024$ ,  $p = 0.022$ ), and  $y_2$  ( $SD_{bc} = 0.030$ ,  $p = 0.001$ ) had significant parameter variation for dataset *DATA1*, while they were not significant for loadings of  $x_2$  and  $y_3$ .

**Listing 6.** Illustrative datasets: Part of the output of `sirt::lsem.bootstrap()` for the illustrative dataset *DATA1*.

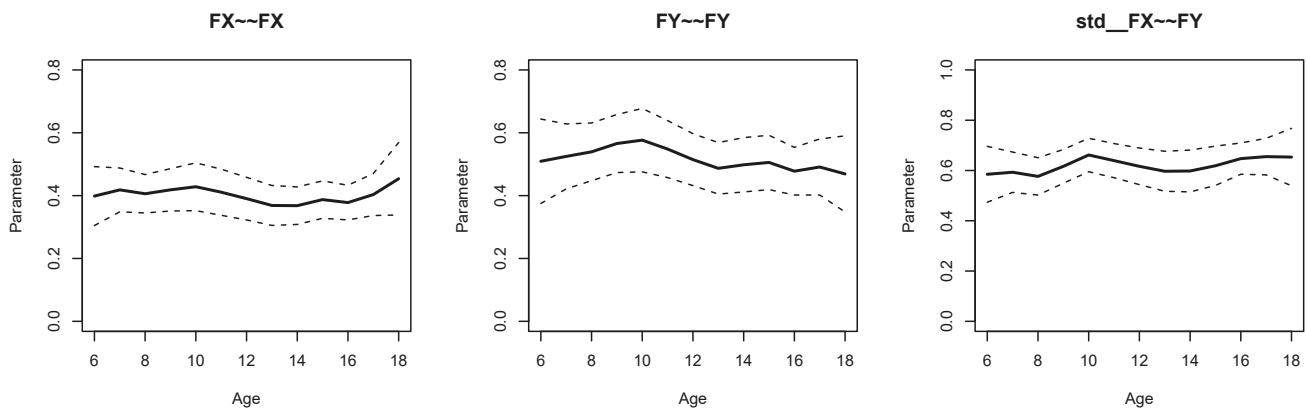
```

1 Global Fit Statistics for Joint Estimation
2
3   stat value value_bc   se
4 1 rmsea 0.017   -0.017 0.007
5 2   cfi 0.999    1.008 0.003
6 3   tli 0.998    1.011 0.004
7 4   gfi 0.988    0.996 0.003
8 5  srmr 0.028    0.016 0.003
9
10 Parameter Estimate Summary
11
12      par parindex      M      SD SD_bc SD_se SD_t SD_p  MAD    Min    Max
13 1      FX=~X1      1  1.000 0.000 0.000 0.000 0.000 0.500 0.000 1.000 1.000
14 2      FX=~X2      2  1.099 0.000 0.000 0.000 0.000 0.500 0.000 1.099 1.099
15 3      FX=~X3      3  0.984 0.000 0.000 0.000 0.000 0.500 0.000 0.984 0.984
16 4      FY=~Y1      4  1.000 0.000 0.000 0.000 0.000 0.500 0.000 1.000 1.000
17 5      FY=~Y2      5  0.866 0.000 0.000 0.000 0.000 0.500 0.000 0.866 0.866
18 6      FY=~Y3      6  0.924 0.000 0.000 0.000 0.000 0.500 0.000 0.924 0.924
19 7      FX~~FX      7  0.396 0.086 0.083 0.011 7.635 0.000 0.064 0.148 0.478
20 8      FY~~FY      8  0.473 0.115 0.111 0.021 5.182 0.000 0.089 0.313 0.760
21 9      FX~~FY      9  0.258 0.077 0.073 0.011 6.538 0.000 0.059 0.083 0.400
22 [...]
23 32 std__FX~~FY     32  0.584 0.072 0.059 0.022 2.736 0.003 0.054 0.387 0.664
24 [...]
25 47 dif__FX=~X1     47  0.996 0.040 0.024 0.012 2.052 0.020 0.032 0.943 1.089
26 48 dif__FX=~X2     48  1.094 0.025 0.000 0.012 0.000 0.500 0.018 1.023 1.127
27 49 dif__FX=~X3     49  0.995 0.043 0.038 0.013 2.812 0.002 0.029 0.946 1.125
28 50 dif__FY=~Y1     50  1.011 0.038 0.024 0.012 2.013 0.022 0.031 0.947 1.076
29 51 dif__FY=~Y2     51  0.860 0.038 0.030 0.010 3.052 0.001 0.032 0.794 0.917
30 52 dif__FY=~Y3     52  0.925 0.024 0.000 0.011 0.000 0.500 0.021 0.882 0.970

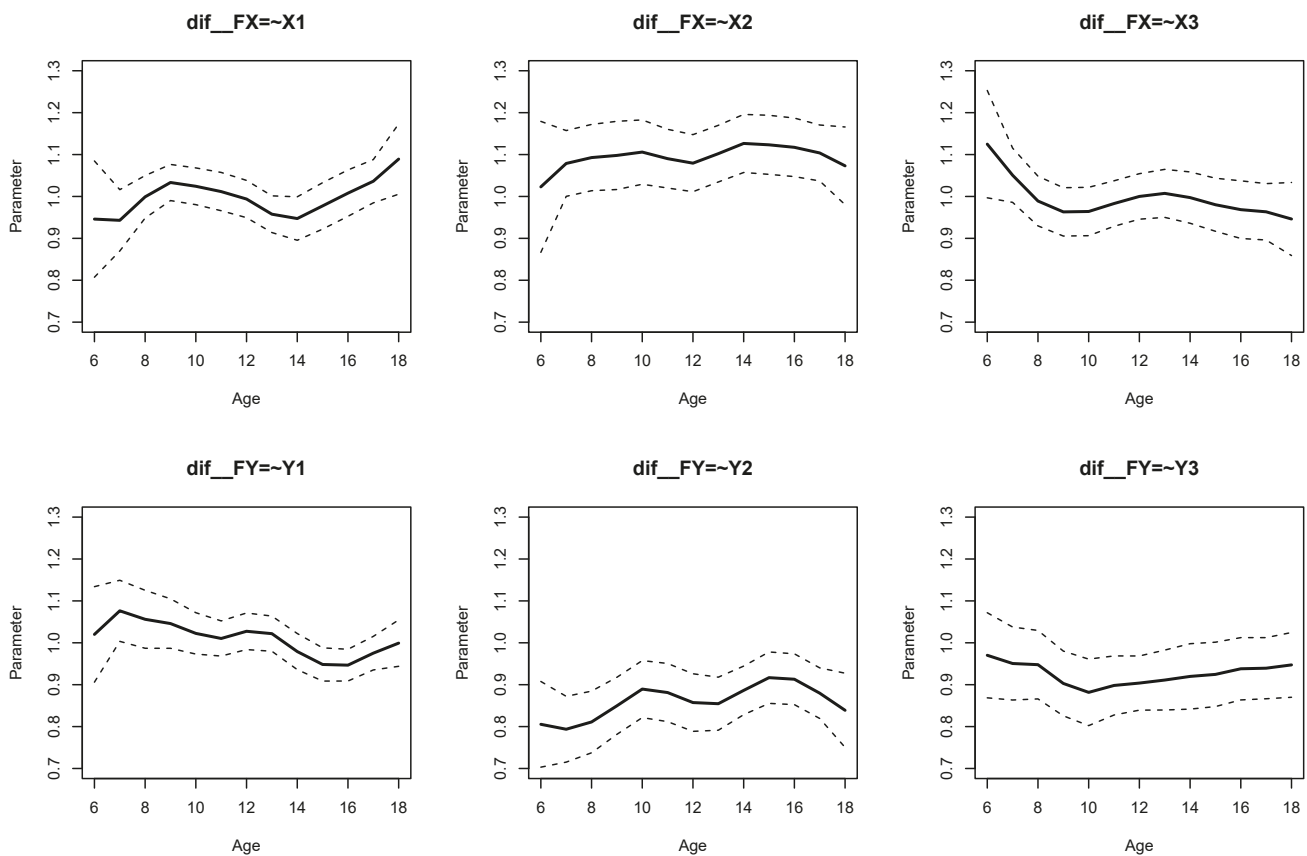
```



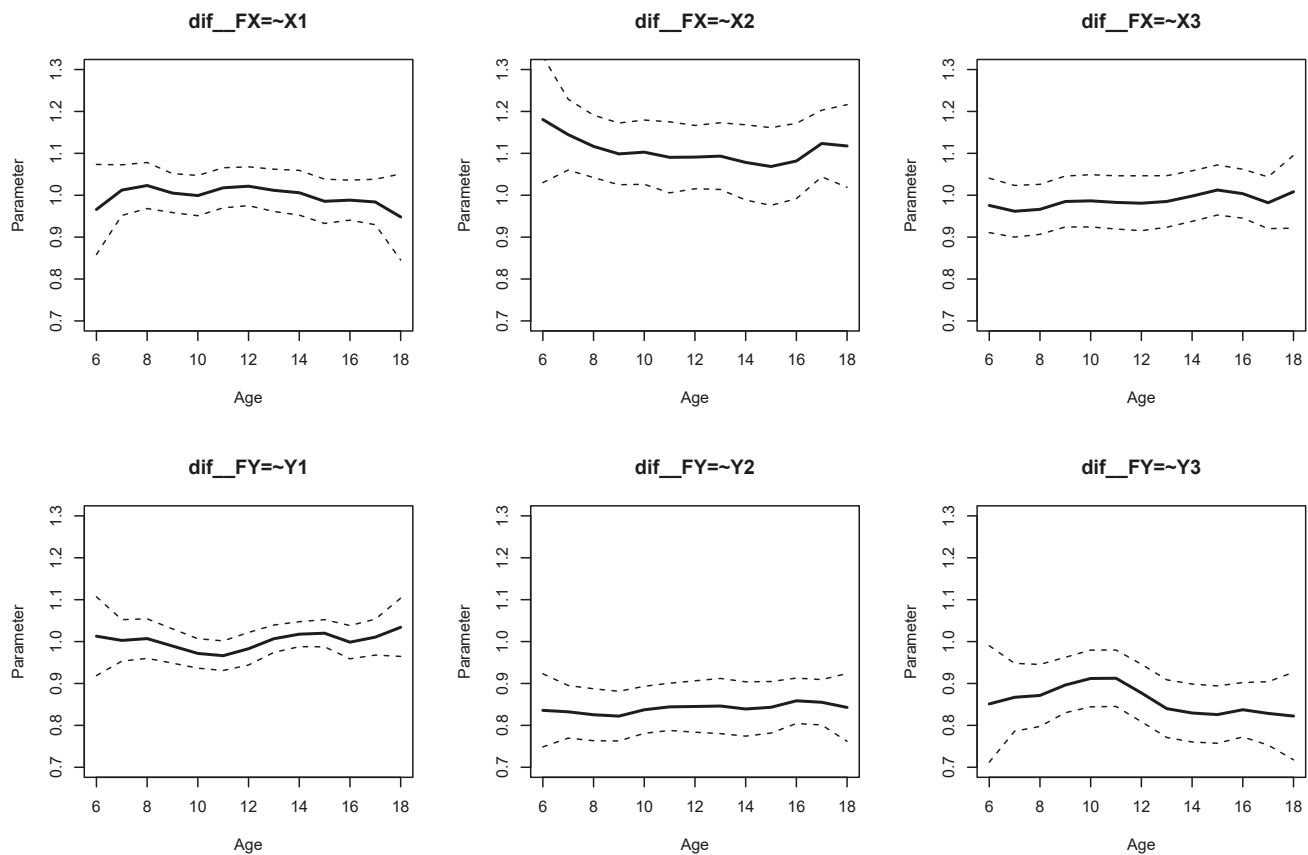
**Figure 1.** Illustrative datasets: Parameter curves for variances of the two factors (i.e.,  $FX \sim FX$  and  $FY \sim FY$ ) and the correlation of the two factors ( $std\_FX \sim FY$ ) for the illustrative dataset *DATA1*.



**Figure 2.** Illustrative datasets: Parameter curves for variances of the two factors (i.e.,  $FX \sim FX$  and  $FY \sim FY$ ) and the correlation of the two factors ( $std\_FX \sim FY$ ) for the illustrative dataset *DATA3*.



**Figure 3.** Illustrative datasets: Parameter curves for DIF effects of factor loadings for the illustrative dataset *DATA1*.



**Figure 4.** Illustrative datasets: Parameter curves for DIF effects of factor loadings for the illustrative dataset *DATA2*.

Finally, part of the R output of `sirt::lsem.bootstrap()` for dataset *DATA3* is displayed in Listing 7. In accordance with the data-generating model, both factor variances, the factor correlation, and the DIF effects for factor loadings did not show significant parameter variation across age.

**Listing 7.** Illustrative datasets: Part of the output of `sirt::lsem.bootstrap()` for the illustrative dataset *DATA3*.

1	Parameter Estimate Summary											
2												
3		par	parindex	M	SD	SD_bc	SD_se	SD_t	SD_p	MAD	Min	Max
4	[...]											
5	7	FX~~FX	7	0.402	0.023	0.000	0.011	0.000	0.500	0.018	0.368	0.454
6	8	FY~~FY	8	0.517	0.032	0.000	0.014	0.000	0.500	0.027	0.469	0.577
7	9	FX~~FY	9	0.282	0.021	0.000	0.009	0.000	0.500	0.017	0.253	0.329
8	[...]											
9	32	std__FX~~FY	32	0.619	0.028	0.000	0.012	0.000	0.500	0.024	0.577	0.662
10	[...]											
11	47	dif__FX=~X1	47	1.000	0.019	0.000	0.011	0.000	0.500	0.016	0.948	1.022
12	48	dif__FX=~X2	48	1.106	0.027	0.000	0.012	0.000	0.500	0.021	1.069	1.175
13	49	dif__FX=~X3	49	0.987	0.014	0.000	0.006	0.000	0.500	0.011	0.965	1.013
14	50	dif__FY=~Y1	50	1.001	0.019	0.000	0.009	0.000	0.500	0.015	0.967	1.037
15	51	dif__FY=~Y2	51	0.842	0.012	0.000	0.007	0.000	0.500	0.010	0.822	0.863
16	52	dif__FY=~Y3	52	0.862	0.034	0.014	0.012	1.161	0.123	0.030	0.805	0.914



Note that a researcher will only have one dataset available for analysis. This section shows that LSEM model parameter output and figures are able to distinguish between situations of noninvariant and invariant model parameters. The standard deviation of a model parameter quantifies the variability of a model parameter across the values of the moderator.

For identification and interpretation reasons, it is useful to specify LSEM models with (some) invariant factor loadings. DIF effects reported in the LSEM output provide a post hoc assessment of the variability of parameter curves across the moderator values if parameter invariance was specified in the LSEM.

## 6. Simulation Study 1: Bias and RMSE

### 6.1. Method

In Simulation Study 1, the bias and the root mean square error (RMSE) of LSEM estimates of parameter curves were investigated. A one-factor model for three indicators,  $x_1$ ,  $x_2$ , and  $x_3$ , with a latent factor variable  $F_X$  was specified. The data-generating model coincided with those from the illustrative datasets presented in Section 5. In contrast to Section 5, we only used the first three observed variables and considered a one-factor instead of a two-factor model in Simulation Study 1.

The population parameters can be found in the directory “POPPARS” at [https://osf.io/puaz9/?view\\_only=63ffb2fd30f5400e89c59d03366bf793](https://osf.io/puaz9/?view_only=63ffb2fd30f5400e89c59d03366bf793) (accessed on 3 June 2023). In this simulation, sample sizes  $N$  were chosen as 250, 500, 1000, 2000, and 4000. Instead of simulating data, random samples without replacement of sample size  $N$  were drawn from population datasets *DATA1* (noninvariant factor loadings, noninvariant factor variances and correlations), resulting in the data-generating model (DGM) DGM1, *DATA2* (invariant factor loadings, noninvariant factor variances and correlations) resulting in DGM2, and *DATA3* (invariant factor loadings, invariant factor variances and correlations), resulting in DGM3. The population datasets that included 130,000 subjects each can be found in the directory “POPDATA” at [https://osf.io/puaz9/?view\\_only=63ffb2fd30f5400e89c59d03366bf793](https://osf.io/puaz9/?view_only=63ffb2fd30f5400e89c59d03366bf793) (accessed on 3 June 2023).

Joint LSEM estimation was carried out using invariant item loadings and bandwidth factor  $h = 1.1, 2$ , and  $3$ , where the bandwidth  $bw$  was defined as  $bw = hN^{-1/5}\hat{\sigma}_A$ . The Gaussian kernel function was used. We also compared the two choices of computing conditional covariances with local smoothing (SM; see (17)) and the weighting approach (16) (no smoothing; NSM). Moreover, we applied LSEM with a quadratic parameter constraint (“quad”) using a bandwidth factor  $h = 1.1$ . A grid of 13 focal points was chosen as  $6, 7, \dots, 18$ .

We investigated the accuracy of the estimated parameter curves of the variance of the latent factor  $F_X$ , the invariant factor loading of the indicator  $X_2$ , and the DIF effect for factor loading of  $X_2$ . Parameter accuracy was assessed by summarizing bias and RMSE of estimated parameter curves across the different age values. The bias of a parameter  $\theta(a_t)$  at a focal point  $a_t$  is given by:

$$Bias(\hat{\theta}(a_t)) = \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r(a_t) - \theta(a_t)) , \quad (33)$$

where  $\hat{\theta}_r(a_t)$  is the parameter estimate of  $\theta(a_t)$  in the  $r$ th replication. The weighted absolute bias can then be defined as:

$$wBias(\hat{\theta}) = \sum_{t=1}^T f(a_t) |Bias(\hat{\theta}(a_t))| , \quad (34)$$

where  $f(a_t)$  denotes the proportion of values of the moderator variable that equal  $a_t$ . The weighted root mean square error (weighted RMSE) is defined as:

$$wRMSE(\hat{\theta}) = \sum_{t=1}^T f(a_t) \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r(a_t) - \theta(a_t))^2}, \quad (35)$$

which is a weighted point-wise RMSE summary statistic.

In total, 2500 replications (i.e., 2500 datasets were generated and analyzed in each condition of the simulation) were conducted. We used the R (R Core Team 2023) software for the entire analysis of the simulation and the *sirt* (Robitzsch 2023b) package for LSEM estimation.

## 6.2. Results

In Table 1, weighted absolute bias and weighted RMSE for the factor variance, the invariant factor loading of  $X_2$ , and the DIF effect of factor loading of  $X_2$  are presented.

It turned out that all three model parameters resulted in unbiased estimation for moderate or large sample sizes. For DGM1 or DGM2, the quadratic parameter constraint introduced some misspecification, which led to slight biases. Moreover, using the local quadratic smoothing approach SM for estimating conditional covariances instead of the weighted approach NM (e.g., no smoothing) resulted in a small error bias. Finally, biases increased with increasing the bandwidth factor  $h$ .

Notably, using local smoothing SM for conditional covariances added variability in terms of RMSE compared to NM. Regarding RMSE, one could conclude that  $h = 2$  seems preferable to  $h = 1.1$  or  $h = 3$  (see also Hildebrandt et al. 2016).

Overall, the findings of Simulation Study 1 demonstrated that joint LSEM estimation resulted in approximately unbiased parameter estimates. The decrease in RMSE values for increasing sample sizes also indicated that parameter estimates are consistent. Notably, the recommendation of using the bandwidth factor  $h = 2$  in pointwise LSEM (Hildebrandt et al. 2016) also transfers to the joint LSEM estimation method.

**Table 1.** Simulation Study 1: Weighted absolute bias and weighted root mean square error (RMSE) for the parameter curve  $\theta(a)$  for different model parameters as a function of sample size  $N$  and three data-generating models DGM1, DGM2 and DGM3.

		Weighted Absolute Bias							Weighted RMSE						
		$h = 1.1$		$h = 2$		$h = 3$			$h = 1.1$		$h = 2$		$h = 3$		
DGM	$N$	SM	NSM	SM	NSM	SM	NSM	Quad	SM	NSM	SM	NSM	SM	NSM	Quad
$Variance\ of\ latent\ factor\ F$															
1	250	0.022	0.022	0.021	0.028	0.025	0.039	0.023	0.088	0.083	0.078	0.076	0.077	0.079	0.076
	500	0.013	0.014	0.015	0.022	0.020	0.033	0.016	0.064	0.061	0.057	0.056	0.056	0.060	0.054
	1000	0.007	0.009	0.010	0.018	0.016	0.029	0.011	0.048	0.046	0.042	0.043	0.042	0.048	0.039
	2000	0.006	0.008	0.008	0.015	0.013	0.025	0.010	0.035	0.034	0.031	0.033	0.032	0.038	0.029
	4000	0.003	0.005	0.005	0.012	0.010	0.021	0.008	0.026	0.026	0.023	0.025	0.023	0.030	0.021
2	250	0.023	0.019	0.021	0.022	0.022	0.030	0.018	0.086	0.080	0.076	0.070	0.073	0.070	0.073
	500	0.013	0.011	0.014	0.017	0.017	0.026	0.011	0.063	0.059	0.055	0.053	0.054	0.054	0.052
	1000	0.008	0.008	0.010	0.015	0.014	0.023	0.008	0.046	0.044	0.040	0.040	0.040	0.042	0.038
	2000	0.005	0.006	0.007	0.012	0.012	0.020	0.006	0.034	0.033	0.030	0.030	0.029	0.032	0.027
	4000	0.003	0.004	0.005	0.009	0.009	0.017	0.005	0.025	0.025	0.022	0.023	0.022	0.026	0.020
3	250	0.018	0.017	0.012	0.010	0.009	0.007	0.017	0.087	0.080	0.076	0.067	0.071	0.061	0.073
	500	0.010	0.009	0.006	0.005	0.004	0.004	0.009	0.063	0.059	0.055	0.049	0.051	0.044	0.052
	1000	0.006	0.006	0.004	0.004	0.003	0.003	0.006	0.047	0.044	0.040	0.036	0.037	0.033	0.038
	2000	0.002	0.002	0.001	0.001	0.001	0.001	0.002	0.034	0.032	0.029	0.026	0.026	0.023	0.026
	4000	0.002	0.002	0.001	0.001	0.001	0.001	0.002	0.025	0.024	0.021	0.020	0.019	0.017	0.019

Table 1. Cont.

DGM	N	Weighted Absolute Bias							Weighted RMSE						
		h = 1.1		h = 2		h = 3		Quad	h = 1.1		h = 2		h = 3		Quad
		SM	NSM	SM	NSM	SM	NSM		SM	NSM	SM	NSM	SM	NSM	
Invariant factor loading of $X_2$															
1	250	0.003	0.005	0.006	0.004	0.008	0.004	0.005	0.089	0.090	0.089	0.089	0.090	0.088	0.089
	500	0.001	0.001	0.000	0.000	0.002	0.000	0.001	0.063	0.063	0.062	0.063	0.063	0.062	0.063
	1000	0.001	0.000	0.001	0.000	0.001	0.000	0.000	0.043	0.043	0.043	0.043	0.044	0.043	0.043
	2000	0.001	0.002	0.000	0.002	0.002	0.001	0.001	0.031	0.031	0.030	0.031	0.031	0.031	0.031
	4000	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.022	0.022	0.022	0.022	0.022	0.022	0.022
2	250	0.005	0.005	0.004	0.004	0.004	0.004	0.005	0.091	0.091	0.091	0.090	0.091	0.090	0.091
	500	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.062	0.062	0.062	0.062	0.062	0.061	0.062
	1000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.045	0.045	0.045	0.045	0.045	0.045	0.045
	2000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.031	0.031	0.031	0.031	0.031	0.031	0.031
	4000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.022	0.022	0.022	0.022	0.022	0.022	0.022
3	250	0.003	0.003	0.002	0.003	0.002	0.003	0.003	0.090	0.090	0.090	0.090	0.090	0.090	0.090
	500	0.003	0.003	0.003	0.003	0.003	0.002	0.003	0.064	0.064	0.063	0.064	0.064	0.064	0.064
	1000	0.000	0.001	0.000	0.001	0.000	0.001	0.001	0.043	0.043	0.043	0.043	0.043	0.044	0.043
	2000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.032	0.032	0.032	0.032	0.032	0.032
	4000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.022	0.022	0.022	0.022	0.022	0.022	0.022
DIF for factor loading of $X_2$															
1	250	0.007	0.007	0.010	0.012	0.013	0.015	0.013	0.123	0.109	0.111	0.098	0.105	0.094	0.123
	500	0.004	0.006	0.008	0.011	0.012	0.014	0.013	0.085	0.079	0.078	0.071	0.075	0.068	0.092
	1000	0.003	0.004	0.006	0.009	0.010	0.013	0.012	0.061	0.057	0.055	0.051	0.053	0.049	0.069
	2000	0.002	0.004	0.005	0.008	0.008	0.012	0.012	0.044	0.041	0.040	0.037	0.038	0.036	0.053
	4000	0.002	0.003	0.004	0.007	0.007	0.010	0.012	0.032	0.031	0.029	0.027	0.028	0.027	0.041
2	250	0.005	0.005	0.005	0.004	0.004	0.004	0.009	0.121	0.110	0.109	0.098	0.105	0.094	0.122
	500	0.002	0.002	0.002	0.002	0.002	0.002	0.008	0.083	0.077	0.075	0.069	0.072	0.065	0.089
	1000	0.001	0.001	0.001	0.001	0.001	0.001	0.008	0.061	0.058	0.055	0.051	0.053	0.048	0.068
	2000	0.001	0.001	0.001	0.001	0.001	0.001	0.007	0.043	0.041	0.039	0.036	0.037	0.034	0.051
	4000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.032	0.030	0.028	0.026	0.027	0.025	0.039
3	250	0.003	0.003	0.003	0.003	0.002	0.003	0.002	0.118	0.109	0.106	0.098	0.102	0.094	0.120
	500	0.003	0.003	0.003	0.003	0.003	0.003	0.002	0.083	0.079	0.076	0.071	0.073	0.067	0.090
	1000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.059	0.056	0.053	0.050	0.050	0.047	0.066
	2000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.044	0.042	0.039	0.037	0.037	0.035	0.051
	4000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.031	0.030	0.028	0.026	0.026	0.025	0.038

## 7. Simulation Study 2: Estimation of Variability of Model Parameters and Statistical Significance Tests

### 7.1. Method

In Simulation Study 2, the bias of standard deviation statistics for parameter variation and the properties of significance tests for parameter variation are investigated. The same three data-generating models DGM1, DGM2, and DGM3 as in Simulation Study 1 (see Section 6.1) were utilized.

The chosen sample sizes in this simulation were  $N = 500, 1000, 2000$ , and  $4000$ . As in Simulation Study 1, samples of sample size  $N$  were drawn without replacement from population datasets *DATA1*, *DATA2*, and *DATA3* for DGM1, DGM2, and DGM3, respectively. The population datasets that included 130,000 subjects can be found in the directory “*POPDATA*” at [https://osf.io/puaz9/?view\\_only=63ffb2fd30f5400e89c59d03366bf793](https://osf.io/puaz9/?view_only=63ffb2fd30f5400e89c59d03366bf793) (accessed on 3 June 2023).

As in Simulation Study 1, a one-factor model with indicators  $x_1$ ,  $x_2$ , and  $x_3$  was specified. Throughout all simulation conditions, a bandwidth factor of  $h = 2$  was chosen. The bias of the two standard deviation estimators  $\widehat{SD}_{\theta(a)}$  and  $\widehat{SD}_{\theta(a),bc}$  defined in (28) and (29) was assessed. Significance testing for parameter variation was based on the standard deviation (see (30)), which uses a normal distribution approximation and the Wald test

(see (31)), which uses a chi-square distribution as a null distribution. Statistical significance tests were performed with significance levels of 0.05 and 0.01. The bias of the standard deviation variability statistics and significance tests of parameter variation was computed for the variance of the latent factor, the three DIF effects of the factor loadings, and the three residual variances.

In total, 2500 replications were conducted in all simulation conditions. The R software (R Core Team 2023) was used for analyzing this simulation study, and the R package sirt (Robitzsch 2023b) was employed for LSEM estimation and significance testing.

## 7.2. Results

In Table 2, the bias of raw and bias-corrected (“bc”) estimates of the standard deviation variability measure  $SD_{\theta(a)}$  are presented. In DGM1, all parameters have nonvanishing  $SD_{\theta(a)}$  values for the population dataset DATA1. In this case, the raw SD estimate showed some slight positive bias for sample sizes  $N = 500$  and 1000. The bias-corrected estimates were generally negatively biased, although the biases were not very large. In DGM2, only the variance of the latent factor  $F$  had a true parameter variation larger than 0. In this situation, raw estimates were approximately unbiased, while the bias-corrected estimates were negatively biased. If there was no true parameter variation, such as for DIF effects or residual variances in DGM2 or all parameters in DGM3, the bias-corrected estimates were less biased than the raw standard deviation estimate.

**Table 2.** Simulation Study 2: Bias of raw and bias-corrected estimators of the standard deviation  $SD_{\theta(a)}$  for the parameter curve  $\theta(a)$  for different model parameters as a function of sample size  $N$  and three data-generating models DGM1, DGM2 and DGM3.

N	DGM 1			DGM 2			DGM 3		
	$SD_{\theta(a)}$			$SD_{\theta(a)}$			$SD_{\theta(a)}$		
	true	raw	bc	true	raw	bc	true	raw	bc
<i>Variance of latent factor F</i>									
500	0.081	−0.002	−0.012	0.054	0.002	−0.013	0	0.035	0.013
1000	0.081	−0.003	−0.009	0.054	−0.001	−0.009	0	0.027	0.010
2000	0.081	−0.003	−0.006	0.054	−0.002	−0.006	0	0.020	0.007
4000	0.081	−0.003	−0.005	0.054	−0.002	−0.004	0	0.015	0.005
<i>DIF for factor loading of <math>X_1</math></i>									
500	0.047	0.007	−0.021	0	0.041	0.013	0	0.039	0.013
1000	0.047	0.001	−0.017	0	0.031	0.010	0	0.029	0.010
2000	0.047	−0.002	−0.012	0	0.023	0.007	0	0.022	0.007
4000	0.047	−0.002	−0.008	0	0.017	0.005	0	0.017	0.005
<i>DIF for factor loading of <math>X_2</math></i>									
500	0.021	0.021	−0.007	0	0.038	0.012	0	0.036	0.012
1000	0.021	0.013	−0.007	0	0.029	0.009	0	0.027	0.009
2000	0.021	0.006	−0.008	0	0.022	0.007	0	0.021	0.007
4000	0.021	0.003	−0.007	0	0.017	0.005	0	0.016	0.005
<i>DIF for factor loading of <math>X_3</math></i>									
500	0.022	0.009	−0.008	0	0.022	0.006	0	0.021	0.006
1000	0.022	0.005	−0.006	0	0.017	0.005	0	0.016	0.005
2000	0.022	0.002	−0.004	0	0.013	0.004	0	0.012	0.004
4000	0.022	0.001	−0.002	0	0.009	0.003	0	0.009	0.003
<i>Residual variance of <math>X_1</math></i>									
500	0.012	0.014	−0.001	0	0.024	0.009	0	0.024	0.009
1000	0.012	0.009	−0.003	0	0.018	0.006	0	0.018	0.006
2000	0.012	0.005	−0.003	0	0.014	0.005	0	0.014	0.005
4000	0.012	0.003	−0.003	0	0.011	0.003	0	0.011	0.003

Table 2. Cont.

N	DGM 1			DGM 2			DGM 3		
	SD $_{\theta(a)}$			SD $_{\theta(a)}$			SD $_{\theta(a)}$		
	true	raw	bc	true	raw	bc	true	raw	bc
<i>Residual variance of X<sub>2</sub></i>									
500	0.007	0.020	0.003	0	0.027	0.010	0	0.027	0.010
1000	0.007	0.014	0.001	0	0.021	0.007	0	0.020	0.007
2000	0.007	0.010	−0.001	0	0.016	0.005	0	0.016	0.005
4000	0.007	0.006	−0.002	0	0.012	0.004	0	0.012	0.004
<i>Residual variance of X<sub>3</sub></i>									
500	0.011	0.009	−0.002	0	0.018	0.007	0	0.018	0.007
1000	0.011	0.006	−0.002	0	0.014	0.005	0	0.014	0.005
2000	0.011	0.003	−0.003	0	0.010	0.003	0	0.011	0.004
4000	0.011	0.002	−0.002	0	0.008	0.002	0	0.008	0.003

true = true value of SD $_{\theta(a)}$  in infinite sample size (i.e., at the population level); raw = raw estimate  $\widehat{SD}_{\theta(a)}$  of SD $_{\theta(a)}$  (see Equation (28)); bc = bias-corrected estimate  $\widehat{SD}_{\theta(a),bc}$  of SD $_{\theta(a)}$  (see Equation (29)).

Overall, one could say that for smaller values of true variability, the positive bias in the raw SD estimate was larger than the underestimation of the bias-corrected SD estimate. An improved SD statistic might be obtained by computing some weighted average of the raw and the bias-corrected estimate.

Table 3 presents type I error and power rates for the different LSEM model parameters. Significance testing based on the SD statistics had inflated type I error rates. If the nominal level was chosen as 1%, the empirical error rate was about 5%. Moreover, the Wald statistic had type I error rates lower than the nominal level in many simulation conditions. Nevertheless, significance testing based on the standard deviation has substantially more statistical power. If a target nominal significance level for the SD test statistic were 5%, it is advised to use a significance level of 0.01.

**Table 3.** Simulation Study 2: Type I and power rates for the significance test for variability in a parameter curve  $\theta(a)$  for the two test statistics based on SD $_{\theta(a)}$  (SD) and the Wald test (WA) as a function of sample size  $N$  and three data-generating models DGM1, DGM2 and DGM3.

N	DGM1				DGM2				DGM3			
	SD5	WA5	SD1	WA1	SD5	WA5	SD1	WA1	SD5	WA5	SD1	WA1
<i>Variance of latent factor F</i>												
500	92.4	46.9	79.8	29.6	66.0	17.8	44.2	8.6	16.3	1.6	4.9	0.5
1000	99.7	88.1	98.5	75.4	90.0	45.7	76.1	26.5	17.4	2.8	5.9	0.8
2000	100.0	99.8	100.0	99.1	99.1	83.0	97.0	65.8	17.7	3.6	5.8	0.9
4000	100.0	100.0	100.0	100.0	100.0	99.5	100.0	98.1	18.3	6.4	7.7	2.0
<i>DIF for factor loading of X<sub>1</sub></i>												
500	23.7	1.3	8.2	0.4	12.9	0.3	3.4	0.1	13.2	0.4	3.4	0.0
1000	46.1	6.0	22.8	1.2	14.8	0.8	5.2	0.2	14.7	0.5	5.1	0.1
2000	75.2	25.2	52.1	10.7	14.4	1.6	4.9	0.3	15.9	1.7	5.5	0.2
4000	96.2	70.6	89.6	48.7	17.5	4.0	6.1	1.1	17.9	4.1	7.1	1.1
<i>DIF for factor loading of X<sub>2</sub></i>												
500	12.3	0.4	3.3	0.1	12.4	0.3	3.5	0.0	12.3	0.3	3.3	0.0
1000	21.7	2.2	8.6	0.5	13.1	0.7	4.1	0.1	13.9	0.8	4.7	0.1
2000	31.2	5.3	14.9	1.5	16.7	1.8	5.8	0.4	16.3	1.4	5.3	0.4
4000	52.4	19.1	32.1	8.2	16.2	3.6	6.9	0.7	18.1	4.2	7.1	1.3



Table 3. Cont.

N	DGM1				DGM2				DGM3			
	SD5	WA5	SD1	WA1	SD5	WA5	SD1	WA1	SD5	WA5	SD1	WA1
<i>DIF for factor loading of <math>X_3</math></i>												
500	18.4	0.4	5.4	0.1	7.5	0.2	1.6	0.0	8.3	0.1	1.8	0.0
1000	38.6	2.0	16.5	0.4	10.5	0.4	2.5	0.0	12.1	0.3	2.9	0.0
2000	66.1	13.7	42.5	4.6	14.2	0.8	4.6	0.0	13.8	1.0	4.6	0.2
4000	92.3	52.0	81.0	28.8	16.1	2.3	5.8	0.5	17.7	2.8	6.2	0.4
<i>Residual variance of <math>X_1</math></i>												
500	20.7	2.3	7.5	0.9	18.2	2.2	6.8	0.5	16.6	2.0	6.2	0.7
1000	25.3	4.0	10.4	1.6	16.6	2.8	5.6	0.9	17.8	3.1	6.5	0.8
2000	34.7	8.4	17.4	2.6	18.3	4.4	6.4	1.2	18.0	4.3	7.1	1.0
4000	49.1	17.7	29.1	7.1	17.3	5.7	7.0	1.5	19.0	6.0	7.2	1.6
<i>Residual variance of <math>X_2</math></i>												
500	16.9	1.9	5.9	0.4	17.7	1.9	6.4	0.4	17.1	1.8	6.3	0.6
1000	18.7	2.5	7.0	0.7	17.6	2.8	6.2	0.7	17.8	2.8	5.7	0.9
2000	22.1	4.8	8.4	1.3	18.5	4.4	7.1	1.3	16.7	3.6	5.9	1.0
4000	29.6	8.8	13.3	3.0	18.0	5.9	7.4	1.6	19.0	6.5	7.1	1.6
<i>Residual variance of <math>X_3</math></i>												
500	25.4	2.9	11.2	0.8	16.7	1.4	5.6	0.5	17.7	2.0	6.1	0.5
1000	34.1	5.5	15.9	1.5	17.9	2.7	6.2	0.7	18.1	2.7	6.4	0.8
2000	47.3	12.0	26.6	4.8	17.2	3.3	6.4	0.8	18.3	4.0	6.7	1.1
4000	68.6	26.3	47.4	12.7	17.9	5.3	7.0	1.4	19.2	6.5	7.7	1.7

Note. SD5 = test statistic based on bias-corrected  $SD_{\theta(a)}$  estimate at 5% confidence level; WA5 = Wald test statistic at 5% confidence level; SD1 = test statistic based on bias-corrected  $SD_{\theta(a)}$  estimate at 1% confidence level;

WA1 = Wald test statistic at 1% confidence level; Cells with yellow-gray colored background correspond to type I error rates, while cells with white background color correspond to power rates.

## 8. Empirical Example: A Reanalysis of SON-R

### 8.1. Data

According to the age differentiation hypothesis, cognitive abilities become more differentiated with increasing age during childhood. Hülür et al. (2011) used data from the German standardization of the SON-R 2<sup>1</sup>/<sub>2</sub>–7 intelligence test to examine age-related differentiation of cognitive abilities from age 2<sup>1</sup>/<sub>2</sub> to age 7. The SON-R 2<sup>1</sup>/<sub>2</sub>–7 intelligence test is a nonverbal intelligence test for children and consists of six indicators (i.e., six subtests). The SON-R 2<sup>1</sup>/<sub>2</sub>–7 test contains two subscales measured by three indicators each. The performance subscale (with factor  $F_p$ ) contains indicators mosaics ( $p_1$ ), puzzles ( $p_2$ ), and patterns ( $p_3$ ). The reasoning subscale (with factor  $F_r$ ) contains the indicators categories ( $r_1$ ), analogies ( $r_2$ ), and situations ( $r_3$ ).

Unfortunately, the SON-R dataset is not publicly available, and the authors of this paper cannot publicly share the dataset on the internet. To replicate the LSEM analysis of this example, we generated a synthetic dataset of the SON-R 2<sup>1</sup>/<sub>2</sub>–7 data based on the original dataset used in Hülür et al. (2011). The same sample size of  $N = 1027$  children was simulated. In the synthetic data generation, we relied on a recently proposed method by Jiang et al. (2021) (see also Grund et al. 2022, Nowok et al. 2016 or Reiter 2023) that combines the distinct approaches of simulating a dataset based on a known distribution and the approach of adding noise to original data to prevent data disclosure or person identification. The noisy versions of the original dataset were simulated with a reliability of 0.95 (Grund et al. 2022), and quadratic relations among variables were allowed. The data synthesis model was separately carried out in 18 groups of children (i.e., in 9 age groups for male and female children, respectively). The values of the age and gender variables were held fixed in the analysis meaning that these demographic variables had the same distribution in the synthetic data as in the original data. In total, 50.8% of the children in the sample was male. The synthetic data and syntax for synthetic data generation can be found

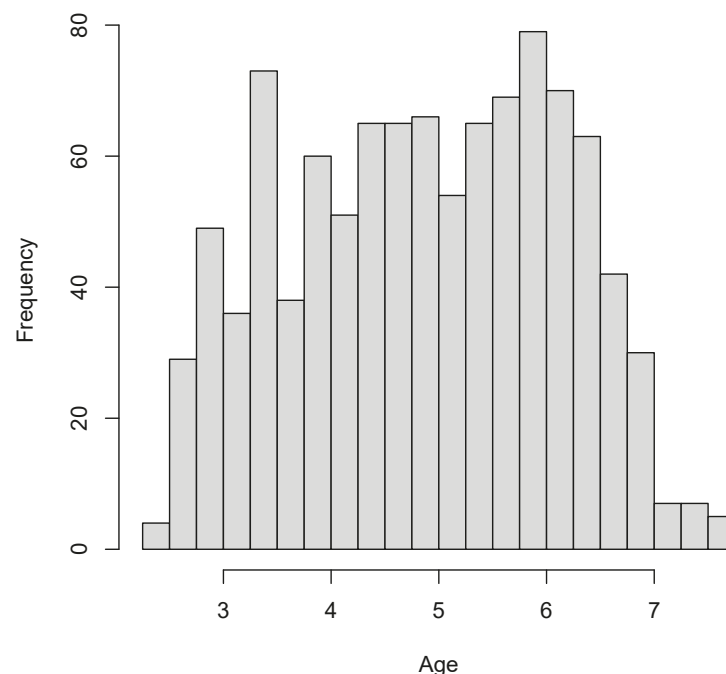
in the directory “SON-R” at [https://osf.io/puaz9/?view\\_only=63ffb2fd30f5400e89c59d03366bf793](https://osf.io/puaz9/?view_only=63ffb2fd30f5400e89c59d03366bf793) (accessed on 3 June 2023).

The indicator variables were linearly transformed such that the mean equaled zero and the standard deviation equaled one for children aged between 4.0 and 6.0 years. This is an arbitrary choice and only affects the scaling of the variables. The assessment of model parameter heterogeneity in the LSEM is independent of this choice. Alternatively, one might also standardize the indicator variables for children in the total sample with ages between 2.5 and 7.5 years.

A two-dimensional CFA model involving the performance and the reasoning factor was specified in an LSEM analysis. The mean structure remained unmodeled because the primary goal of this analysis was to investigate the age differentiation hypothesis. For model identification, the factor loadings were assumed as invariant across age, and the first loading of both scales (i.e., loadings of  $p_1$  and  $r_1$ ) were fixed at one. In accordance with Hildebrandt et al. (2016) and the findings of Simulation Study 1, the bandwidth factor of  $h = 2$  was chosen, resulting in a bandwidth  $bw = 2N^{-1/5}\hat{\sigma}_A$ , where  $\hat{\sigma}_A = 1.23$  is the estimated standard deviation of the age variable. Because the LSEM model involved invariance constraints among parameters, a joint estimation approach was employed. For statistical inference and the test of parameter variation,  $R = 200$  bootstrap samples were drawn. Replication syntax can also be found in the directory “SON-R” at [https://osf.io/puaz9/?view\\_only=63ffb2fd30f5400e89c59d03366bf793](https://osf.io/puaz9/?view_only=63ffb2fd30f5400e89c59d03366bf793) (accessed on 3 June 2023).

## 8.2. Results

Figure 5 displays the histogram of the age variable. The age of children ranged between 2.44 and 7.72 years, with a mean of 4.89 and a standard deviation of 1.23. The histogram indicated that the intended age range between 2.5 and 7 years of the SON-R 2½–7 test was approximately uniformly distributed.



**Figure 5.** SON-R example: Histogram for moderator age.

The estimated LSEM model had an acceptable model fit regarding typical model fit effect sizes. The fit statistics without bias correction were RMSEA = 0.061, CFI = 0.952, TLI = 0.960, GFI = 0.963, and SRMR = 0.055.

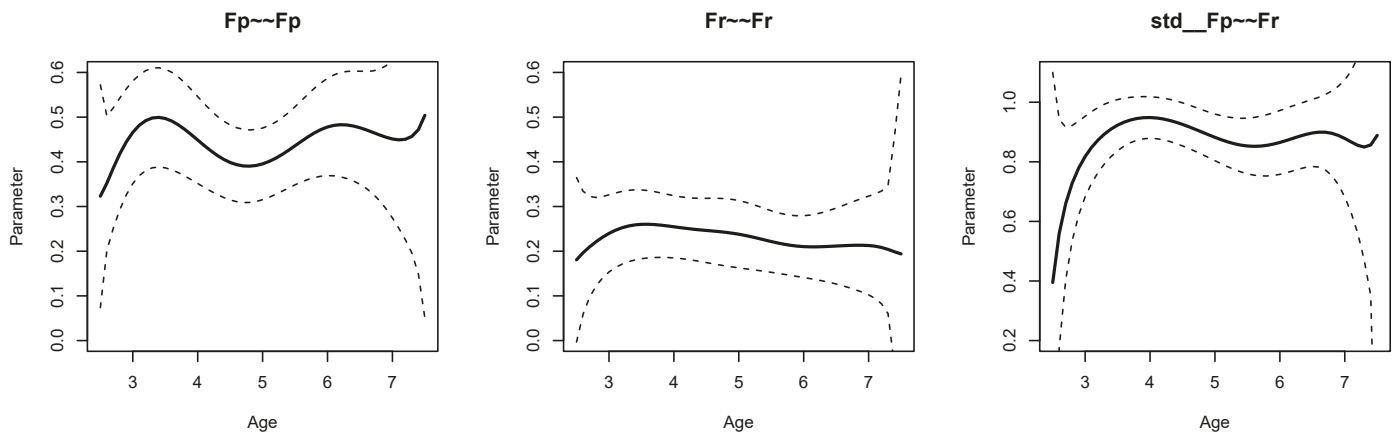
In Listing 8, parts of the LSEM output of `lsem.bootstrap()` are displayed. According to the specified model, the parameter variation (i.e., SD and SD\_bc) for factor loadings (i.e.,  $F_{p \sim p_1, \dots, F_{3 \sim r_3}}$ ) was zero because the parameters were assumed invariant across age.

**Listing 8.** SON-R example: Part of the output of `lsem.bootstrap()` function.

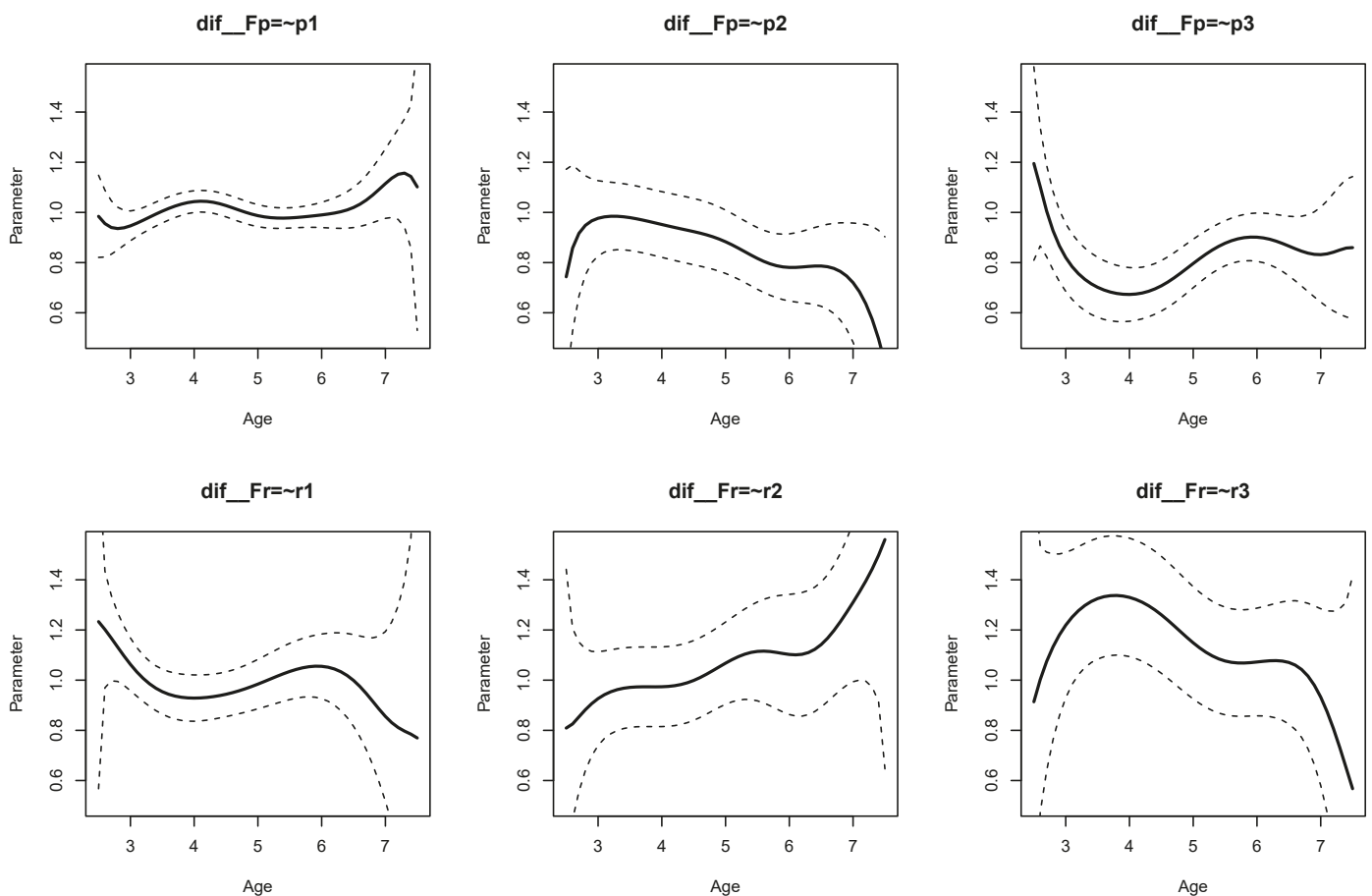
1	Parameter	Estimate	Summary									
2												
3		par	parindex	M	SD	SD_bc	SD_se	SD_t	SD_p	MAD	Min	Max
4	1	Fp=~p1	1	1.000	0.000	0.000	0.000	0.000	0.500	0.000	1.000	1.000
5	2	Fp=~p2	2	0.854	0.000	0.000	0.000	0.000	0.500	0.000	0.854	0.854
6	3	Fp=~p3	3	0.803	0.000	0.000	0.000	0.000	0.500	0.000	0.803	0.803
7	4	Fr=~r1	4	1.000	0.000	0.000	0.000	0.000	0.500	0.000	1.000	1.000
8	5	Fr=~r2	5	1.041	0.000	0.000	0.000	0.000	0.500	0.000	1.041	1.041
9	6	Fr=~r3	6	1.160	0.000	0.000	0.000	0.000	0.500	0.000	1.160	1.160
10	7	Fp~~Fp	7	0.445	0.038	0.000	0.016	0.000	0.500	0.033	0.323	0.504
11	8	Fr~~Fr	8	0.232	0.019	0.000	0.013	0.000	0.500	0.017	0.181	0.260
12	9	Fp~~Fr	9	0.282	0.033	0.017	0.011	1.530	0.063	0.021	0.095	0.334
13	10	p1~~p1	10	0.347	0.082	0.076	0.020	3.777	0.000	0.057	0.032	0.532
14	11	p2~~p2	11	0.498	0.074	0.060	0.021	2.828	0.002	0.064	0.365	0.611
15	12	p3~~p3	12	0.411	0.046	0.030	0.015	2.057	0.020	0.031	0.122	0.573
16	13	r1~~r1	13	0.510	0.120	0.114	0.024	4.734	0.000	0.098	0.172	0.636
17	14	r2~~r2	14	0.510	0.110	0.103	0.024	4.206	0.000	0.096	0.333	0.675
18	15	r3~~r3	15	0.571	0.055	0.019	0.022	0.887	0.188	0.046	0.431	0.677
19	[...]											
20	32	std__Fp~~Fr	32	0.878	0.071	0.019	0.041	0.480	0.316	0.041	0.395	0.949
21	[...]											
22	47	dif__Fp=~p1	47	1.005	0.038	0.013	0.016	0.833	0.202	0.029	0.936	1.157
23	48	dif__Fp=~p2	48	0.868	0.089	0.077	0.029	2.662	0.004	0.077	0.400	0.985
24	49	dif__Fp=~p3	49	0.808	0.097	0.093	0.020	4.749	0.000	0.082	0.673	1.195
25	50	dif__Fr=~r1	50	0.994	0.068	0.000	0.043	0.000	0.500	0.054	0.770	1.233
26	51	dif__Fr=~r2	51	1.053	0.105	0.078	0.043	1.803	0.036	0.082	0.809	1.560
27	52	dif__Fr=~r3	52	1.159	0.126	0.092	0.045	2.056	0.020	0.107	0.567	1.338

The age differentiation hypothesis refers to the variances of the performance scale (i.e.,  $Fp \sim Fp$ ), the variance of the reasoning scale (i.e.,  $Fr \sim Fr$ ), and the correlation of both factors (i.e.,  $std\_Fp \sim Fr$ ). Figure 6 displays the parameter curves with confidence intervals for the two variances and the correlation. From the R output presented in Listing 8, it can be seen that the variances parameter curves did not show significant parameter variation, and the bias-corrected standard deviation estimate  $SD_{bc}$  was 0.000. The correlation between the performance and the reasoning scale was 0.878 on average, with a small bias-corrected standard deviation estimate of 0.019 that turned out to be nonsignificant ( $p = 0.316$ ). Hence, there was no evidence for the age differentiation hypothesis in the SON-R dataset.

Figure 7 displays the parameter curves of the DIF effects of the factor loadings. The corresponding parameters for DIF effects can be found in lines 47 to 52 in Listing 8 (i.e., parameters  $dif\_Fp \sim p1$ , ...,  $dif\_Fr \sim r3$ ). There was substantial parameter variation in terms of the bias-corrected standard deviation  $SD_{bc}$  for the loadings of  $p2$  ( $SD_{bc} = 0.077$ ,  $p = 0.004$ ),  $p3$  ( $SD_{bc} = 0.077$ ,  $p < 0.001$ ),  $r2$  ( $SD_{bc} = 0.078$ ,  $p = 0.036$ ), and  $r3$  ( $SD_{bc} = 0.092$ ,  $p = 0.020$ ).

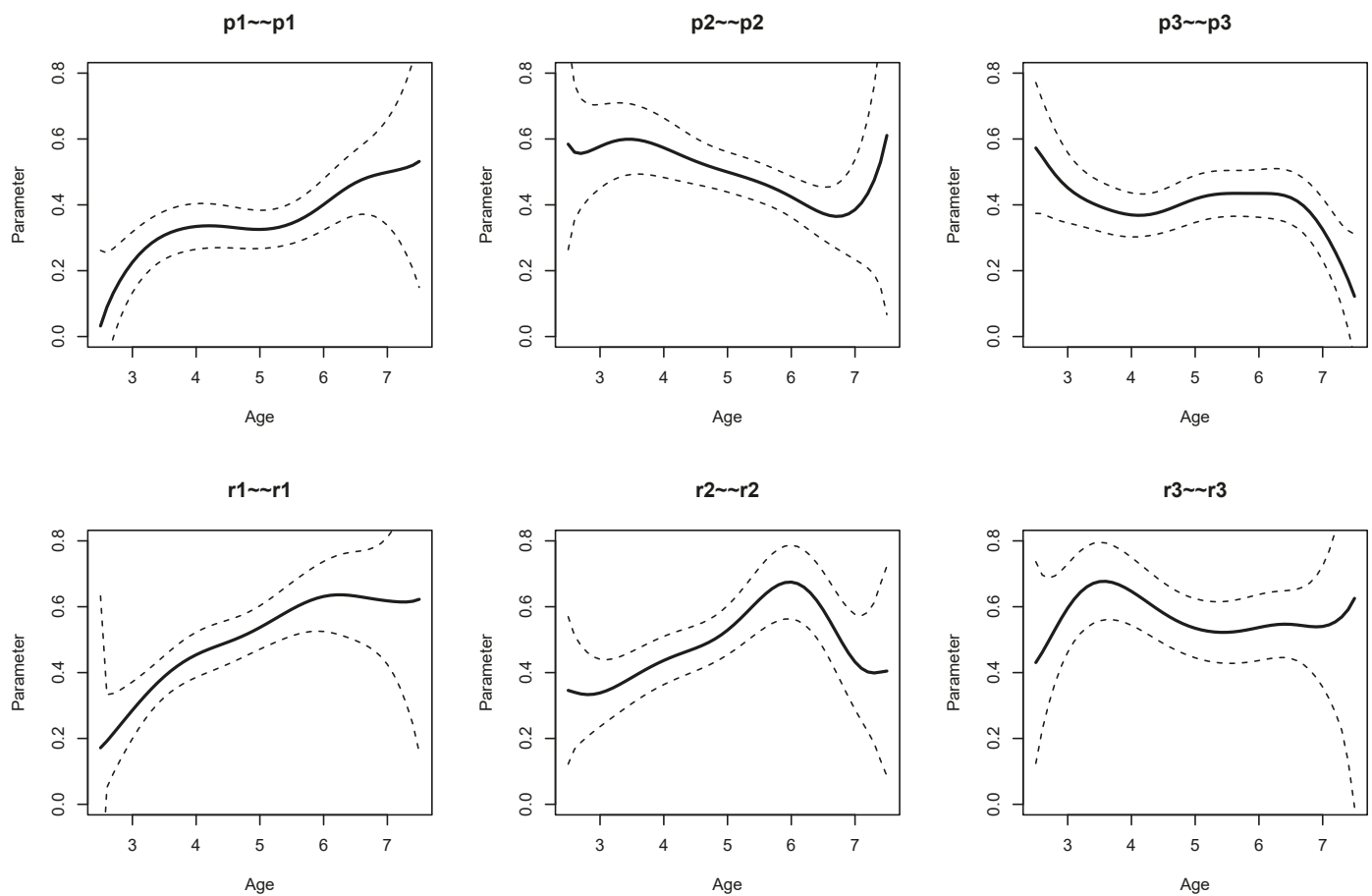


**Figure 6.** SON-R example: Parameter curves for variances of performance (Fp~~Fp) and reasoning (Fr~~Fr) and the correlation of performance and reasoning (std\_Fp~~Fr).



**Figure 7.** SON-R example: Parameter curves for DIF effects of factor loadings for performance (latent variable Fp) and reasoning (latent variable Fr).

Finally, residual variances are displayed in Figure 8. From the results from Listing 8, it is evident that residual variances of p1 ( $SD_{bc} = 0.076$ ,  $p < 0.001$ ), p2 ( $SD_{bc} = 0.060$ ,  $p = 0.002$ ), r1 ( $SD_{bc} = 0.114$ ,  $p < 0.001$ ), and r2 ( $SD_{bc} = 0.103$ ,  $p < 0.001$ ) were statistically significant at the 0.01 significance level.



**Figure 8.** SON-R example: Parameter curves for residual variances.

Note that [Hülür et al. \(2011\)](#) used a pointwise LSEM approach instead a joint LSEM estimation approach. The identification of parameters in the covariance structure of factors was achieved in [Hülür et al. \(2011\)](#) by the constraint that the pointwise average of factor loadings equaled 1. Due to the different estimation approaches, it is expected that there are slight differences between our joint LSEM estimation approach and the original analysis in [Hülür et al. \(2011\)](#). The parameter curve of the correlation between the performance and the reasoning factors was similar in both analyses, with the exception that the factor correlation for small age values was much lower in the joint estimation approach, as displayed in [Figure 6](#).

An anonymous reviewer wondered whether the factor correlation could be meaningfully interpreted if factor loadings did not show invariance across the moderator values. We argued elsewhere that measurement invariance would be a helpful but not a necessary condition for a meaningful interpretation of a factor correlation or a factor variance (see [Robitzsch and Lüdtke 2023](#)). In fact, a violation of measurement invariance only implies that results would change if a subset of indicators was used in the factor model. Because the SON-R instrument is held fixed in test administration and statistical analysis, this property of item selection invariance is not required. Of course, any identification constraint on factor loadings must be imposed to identify a factor correlation. The choice of identification constraint is somehow arbitrary. It could be invariance of all factor loadings, invariance of loadings of a subset of indicators, or a pointwise constraint of the average loadings (i.e., the average loading should be 1 for all indicators of a factor).



## 9. Discussion

In this article, we discussed the implementation of LSEM in the R package *sirt*. Joint LSEM estimation and two different significance tests for a test of parameter variation were introduced and evaluated through two simulation studies.

Simulation Study 1 demonstrated that the joint LSEM estimation method can be successfully applied to structural equation models whose parameters vary across different values of the moderator variable. It turned out that the bandwidth factor  $h = 2$  can generally be recommended as a default choice. Notably, LSEM model parameters can be quite variable for small ( $N = 250$ ) or moderate sample sizes ( $N = 500$ ). In Simulation Study 2, two significance testing approaches for constant parameter curves were investigated: a test statistic based on the standard deviation of a parameter curve and a Wald-type test statistic. Both testing approaches rely on bootstrap samples for statistical inference. The standard-deviation-based test statistics had a higher power than the Wald test-type test statistic, but also came with an inflated type-I error rate. It is recommended to use the significance test based on the standard deviation with a significance level of 1% if a nominal significance level of 5% is required.

The application of LSEM in applied research can be regarded more as an exploratory than a confirmatory statistical method (Jacobucci 2022). Functional forms of parameter curves obtained with LSEM can be validated in other samples or future studies with more confirmatory approaches, such as moderated nonlinear factor analysis. We would like to emphasize that sufficiently large sample sizes are required in LSEM in order to allow a reliable interpretation of the obtained nonlinear parameter curves. Moreover, the true variability in parameter curves must be sufficiently large to have enough power to statistically detect the significant parameter variability. A statistical significance test on parameter curve regression coefficients in a moderated nonlinear factor analysis might have more power than a test based on the nonparametric LSEM method. Finally, moderated nonlinear factor analysis, if estimated by maximum likelihood, allows likelihood ratio tests for testing among nested models or using information criteria for model comparisons.

In this article, the moderator variable was exclusively age and a bounded variable. There might be applications in which the moderator differs from age, such as unbounded self-concept factor variables or ability values obtained from item response models (Basarkod et al. 2023). Because the metric of such variables is often arbitrary, it is advised to transform such moderators into a bounded metric. For example, the percentage ranks of an unbounded moderator variable could be utilized to obtain a bounded moderator variable.

If the moderator variable is an error-prone variable such as a factor variable or a scale score, an expected a posteriori (EAP) factor score estimate can be used as a moderator to obtain unbiased estimates of LSEM model parameters (Bartholomew et al. 2011).

As explained in Section 4, datasets with missing values should either be handled with pairwise deletion methods for computing sufficient statistics (i.e., the conditional covariance matrices) in LSEM or should be multiply imputed. The imputation model should be flexibly specified to represent the complex associations modeled with LSEM. For example, the moderator variable could be discretized into 5 or 10 distinct groups, and the resulting datasets should be separately imputed in the separate subdatasets. Statistical inference should be carried out that involves the multiply imputed datasets (Little and Rubin 2002).

Finally, we only discussed LSEM in the case of one moderator variable. With more than one moderator variable (Hartung et al. 2018), moderated nonlinear factor analysis might be easier to estimate because multivariate kernel functions for LSEM are difficult to estimate with sparse data.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Datasets and R code is available as supplementary material at [https://osf.io/puaz9/?view\\_only=63ffb2fd30f5400e89c59d03366bf793](https://osf.io/puaz9/?view_only=63ffb2fd30f5400e89c59d03366bf793) (accessed on 3 June 2023).

**Conflicts of Interest:** The author declares 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
LSEM	local structural equation model
ML	maximum likelihood
MNFA	moderated nonlinear factor analysis
SEM	structural equation model

## References

- Allemand, Mathias, Gabriel Olaru, and Patrick L. Hill. 2021. Age-related psychometrics and differences in gratitude and future time perspective across adulthood. *Personality and Individual Differences* 182: 111086. [\[CrossRef\]](#)
- Allemand, Mathias, Gabriel Olaru, and Patrick L. Hill. 2022. Gratitude and future time perspective during the COVID-19 pandemic: Effects of age and virus worry. *The Journal of Positive Psychology* 17: 819–31. [\[CrossRef\]](#)
- Arnold, Manuel, Andreas M. Brandmaier, and Manuel C. Voelkle. 2021. Predicting differences in model parameters with individual parameter contribution regression using the R package ipcr. *Psych* 3: 360–85. [\[CrossRef\]](#)
- Arnold, Manuel, Daniel L. Oberski, Andreas M. Brandmaier, and Manuel C. Voelkle. 2020. Identifying heterogeneity in dynamic panel models with individual parameter contribution regression. *Structural Equation Modeling: A Multidisciplinary Journal* 27: 613–28. [\[CrossRef\]](#)
- Bartholomew, David J., Martin Knott, and Irini Moustaki. 2011. *Latent Variable Models and Factor Analysis: A Unified Approach*. New York: Wiley. [\[CrossRef\]](#)
- Basarkod, Geetanjali, Herbert W. Marsh, Baljinder K. Sahdra, Philip D. Parker, Jiesi Guo, Theresa Dicke, and Oliver Lüdtke. 2023. The dimensionality of reading self-concept: Examining its stability using local structural equation models. *Assessment* 30: 873–90. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bates, Timothy C., Hermine Maes, and Michael C. Neale. 2019. umx: Twin and path-based structural equation modeling in R. *Twin Research and Human Genetics* 22: 27–41. [\[CrossRef\]](#)
- Bauer, Daniel J. 2017. A more general model for testing measurement invariance and differential item functioning. *Psychological Methods* 22: 507–26. [\[CrossRef\]](#)
- Bentler, Peter M., and Ke-Hai Yuan. 2011. Positive definiteness via off-diagonal scaling of a symmetric indefinite matrix. *Psychometrika* 76: 119–23. [\[CrossRef\]](#)
- Boker, Steven, Michael Neale, Hermine Maes, Michael Wilde, Michael Spiegel, Timothy Brick, Jeffrey Spies, Ryne Estabrook, Sarah Kenny, Timothy Bates, and et al. 2011. OpenMx: An open source extended structural equation modeling framework. *Psychometrika* 76: 306–17. [\[CrossRef\]](#)
- Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley. [\[CrossRef\]](#)
- Bollen, Kenneth A., and Walter R. Davis. 2009. Two rules of identification for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal* 16: 523–36. [\[CrossRef\]](#)
- Bolsinova, Maria, and Dylan Molenaar. 2018. Modeling nonlinear conditional dependence between response time and accuracy. *Frontiers in Psychology* 9: 1525. [\[CrossRef\]](#)
- Bolsinova, Maria, and Dylan Molenaar. 2019. Nonlinear indicator-level moderation in latent variable models. *Multivariate Behavioral Research* 54: 62–84. [\[CrossRef\]](#) [\[PubMed\]](#)
- Boos, Dennis D., and Leonard A. Stefanski. 2013. *Essential Statistical Inference*. New York: Springer. [\[CrossRef\]](#)
- Brandt, Naemi D., Johanna Drewelies, Sherry L. Willis, K. Warner Schaie, Nilam Ram, Denis Gerstorf, and Jenny Wagner. 2022. Beyond big five trait domains: Stability and change in personality facets across midlife and old age. *Journal of Personality*, Epub ahead of print. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bratt, Christopher, Dominic Abrams, Hannah J. Swift, Christin-Melanie Vauclair, and Sibila Marques. 2018. Perceived age discrimination across age in Europe: From an ageing society to a society for all ages. *Developmental Psychology* 54: 167–80. [\[CrossRef\]](#) [\[PubMed\]](#)
- Breit, Moritz, Julian Preuß, Vsevolod Scherrer, Tobias Moors, and Franzis Preckel. 2023. Relationship between creativity and intelligence: A multimethod investigation of alternative theoretical assumptions in two samples of secondary school students. *Gifted Child Quarterly* 67: 95–109. [\[CrossRef\]](#)

- Breit, Moritz, Martin Brunner, and Franzis Preckel. 2020. General intelligence and specific cognitive abilities in adolescence: Tests of age differentiation, ability differentiation, and their interaction in two large samples. *Developmental Psychology* 56: 364–84. [CrossRef]
- Breit, Moritz, Martin Brunner, and Franzis Preckel. 2021. Age and ability differentiation in children: A review and empirical investigation. *Developmental Psychology* 57: 325–46. [CrossRef]
- Breit, Moritz, Martin Brunner, Dylan Molenaar, and Franzis Preckel. 2022. Differentiation hypotheses of intelligence: A systematic review of the empirical evidence and an agenda for future research. *Psychological Bulletin* 148: 518–54. [CrossRef]
- Briley, Daniel A., K. Paige Harden, and Elliot M. Tucker-Drob. 2015b. Genotype  $\times$  cohort interaction on completed fertility and age at first birth. *Behavior Genetics* 45: 71–83. [CrossRef]
- Briley, Daniel A., K. Paige Harden, Timothy C. Bates, and Elliot M. Tucker-Drob. 2015a. Nonparametric estimates of gene  $\times$  environment interaction using local structural equation modeling. *Behavior Genetics* 45: 581–96. [CrossRef]
- Browne, Michael W. 1974. Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal* 8: 1–24. Available online: <https://bit.ly/3yviejm> (accessed on 3 June 2023). [CrossRef]
- Browne, Michael W., and Gerhard Arminger. 1995. Specification and estimation of mean-and covariance-structure models. In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. Edited by Gerhard Arminger, Clifford C. Clogg and Michael E. Sobel. Boston: Springer, pp. 185–249. [CrossRef]
- Cox, Simon R., Stuart J. Ritchie, Elliot M. Tucker-Drob, David C. Liewald, Saskia P. Hagenaars, Gail Davies, Joanna M. Wardlaw, Catharine R. Gale, Mark E. Bastin, and Ian J. Deary. 2016. Ageing and brain white matter structure in 3513 UK Biobank participants. *Nature Communications* 7: 1–13. [CrossRef]
- Curran, Patrick J., James S. McGinley, Daniel J. Bauer, Andrea M. Hussong, Alison Burns, Laurie Chassin, Kenneth Sher, and Robert Zucker. 2014. A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate Behavioral Research* 49: 214–31. [CrossRef] [PubMed]
- de Mooij, Susanne M. M., Richard N. A. Henson, Lourens J. Waldorp, and Rogier A. Kievit. 2018. Age differentiation within gray matter, white matter, and between memory and white matter in an adult life span cohort. *Journal of Neuroscience* 38: 5826–36. [CrossRef]
- Dong, Yixiao, and Denis Dumas. 2020. Are personality measures valid for different populations? A systematic review of measurement invariance across cultures, gender, and age. *Personality and Individual Differences* 160: 109956. [CrossRef]
- Dutton, Edward, and Emil Kirkegaard. 2022. The negative religiousness-IQ nexus is a Jensen effect on individual-level data: A refutation of Dutton et al.'s "The myth of the stupid believer". *Journal of Religion and Health* 61: 3253–75. [CrossRef] [PubMed]
- Efron, Bradley, and Robert J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Boca Raton: CRC Press. [CrossRef]
- Entringer, Theresa M., and Samuel D. Gosling. 2022. Loneliness during a nationwide lockdown and the moderating effect of extroversion. *Social Psychological and Personality Science* 13: 769–80. [CrossRef]
- Fox, John. 2016. *Applied Regression Analysis and Generalized Linear Models*. Thousand Oaks: Sage. Available online: <https://bit.ly/38XUSX1> (accessed on 3 June 2023).
- Gana, Kamel, Nedjem Eddine Boudouda, Thomas Salanova, and Guillaume Broc. 2023. What do EURO-D scores capture? Disentangling trait and state variances in depression symptoms across the adult life span in nine European nations. *Assessment*, Epub ahead of print. [CrossRef]
- Gnams, Timo. 2013. The elusive general factor of personality: The acquaintance effect. *European Journal of Personality* 27: 507–20. [CrossRef]
- Gnams, Timo, and Ulrich Schroeders. 2020. Cognitive abilities explain wording effects in the Rosenberg self-esteem scale. *Assessment* 27: 404–18. [CrossRef]
- Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984. Pseudo maximum likelihood methods: Theory. *Econometrica* 52: 681–700. [CrossRef]
- Grund, Simon, Oliver Lüdtke, and Alexander Robitzsch. 2022. Using synthetic data to improve the reproducibility of statistical results in psychological research. *Psychological Methods*, Epub ahead of print. [CrossRef] [PubMed]
- Han, Kyunghee, Stephen M. Colarelli, and Nathan C. Weed. 2019. Methodological and statistical advances in the consideration of cultural diversity in assessment: A critical review of group classification and measurement invariance testing. *Psychological Assessment* 31: 1481–96. [CrossRef]
- Hartung, Johanna, Laura E. Engelhardt, Megan L. Thibodeaux, K. Paige Harden, and Elliot M. Tucker-Drob. 2020. Developmental transformations in the structure of executive functions. *Journal of Experimental Child Psychology* 189: 104681. [CrossRef] [PubMed]
- Hartung, Johanna, Martina Bader, Morten Moshagen, and Oliver Wilhelm. 2022. Age and gender differences in socially aversive ("dark") personality traits. *European Journal of Personality* 36: 3–23. [CrossRef]
- Hartung, Johanna, Philipp Doebler, Ulrich Schroeders, and Oliver Wilhelm. 2018. Dedifferentiation and differentiation of intelligence in adults across age and years of education. *Intelligence* 69: 37–49. [CrossRef]
- Hartung, Johanna, Sandy S. Spormann, Morten Moshagen, and Oliver Wilhelm. 2021. Structural differences in life satisfaction in a US adult sample across age. *Journal of Personality* 89: 1232–51. [CrossRef]
- Hildebrandt, Andrea, Oliver Lüdtke, Alexander Robitzsch, Christopher Sommer, and Oliver Wilhelm. 2016. Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research* 51: 257–78. [CrossRef]

- Hildebrandt, Andrea, Oliver Wilhelm, and Alexander Robitzsch. 2009. Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology* 16: 87–102.
- Hildebrandt, Andrea, Oliver Wilhelm, Grit Herzmann, and Werner Sommer. 2013. Face and object cognition across adult age. *Psychology and Aging* 28: 243–48. [CrossRef] [PubMed]
- Hildebrandt, Andrea, Werner Sommer, Grit Herzmann, and Oliver Wilhelm. 2010. Structural invariance and age-related performance differences in face cognition. *Psychology and Aging* 25: 794–810. [CrossRef]
- Holland, Paul W., and Howard Wainer, eds. 1993. *Differential Item Functioning: Theory and Practice*. Hillsdale: Lawrence Erlbaum. [CrossRef]
- Hülür, Gizem, Oliver Wilhelm, and Alexander Robitzsch. 2011. Intelligence differentiation in early childhood. *Journal of Individual Differences* 32: 170–79. [CrossRef]
- Huth, Karoline B. S., Lourens J. Waldorp, Judy Luigjes, Anneke E. Goudriaan, Ruth J. van Holst, and Marten Marsman. 2022. A note on the structural change test in highly parameterized psychometric models. *Psychometrika* 87: 1064–80. [CrossRef]
- Jacobucci, Ross. 2022. A critique of using the labels confirmatory and exploratory in modern psychological research. *Frontiers in Psychology* 13: 1020770. [CrossRef] [PubMed]
- Jiang, Bei, Adrian E. Raftery, Russell J. Steele, and Naisyin Wang. 2021. Balancing inferential integrity and disclosure risk via model targeted masking and multiple imputation. *Journal of the American Statistical Association* 117: 52–66. [CrossRef]
- Jokić-Begić, Nataša, Una Mikac, Doris Čuržik, and Claire Sangster Jokić. 2019. The development and validation of the short cyberchondria scale (SCS). *Journal of Psychopathology and Behavioral Assessment* 41: 662–76. [CrossRef]
- Jöreskog, Karl G., Ulf H. Olsson, and Fan Y. Wallentin. 2016. *Multivariate Analysis with LISREL*. Basel: Springer. [CrossRef]
- Kaltwasser, Laura, Una Mikac, Vesna Buško, and Andrea Hildebrandt. 2017. No robust association between static markers of testosterone and facets of socio-economic decision making. *Frontiers in Behavioral Neuroscience* 11: 250. [CrossRef]
- Klieme, Katrin E., and Florian Schmidt-Borcherding. 2023. Lacking measurement invariance in research self-efficacy: Bug or feature? *Frontiers in Education* 8: 1092714. [CrossRef]
- Kolbe, Laura, Dylan Molenaar, Suzanne Jak, and Terrence D. Jorgensen. 2022. Assessing measurement invariance with moderated nonlinear factor analysis using the R package OpenMx. *Psychological Methods*, Epub ahead of print. [CrossRef]
- Kolenikov, Stanislav. 2010. Resampling variance estimation for complex survey data. *The Stata Journal* 10: 165–99. [CrossRef]
- Kolenikov, Stanislav. 2011. Biases of parameter estimates in misspecified structural equation models. *Sociological Methodology* 41: 119–57. [CrossRef]
- Krasko, Julia, and Till Kaiser. 2023. Die Dunkle Triade in einer deutschen repräsentativen Stichprobe [The dark triad in a German representative sample]. *Diagnostica* 69: 1–13. [CrossRef]
- Leitgöb, Heinz, Daniel Seddig, Tihomir Asparouhov, Dorothée Behr, Eldad Davidov, Kim De Roover, Suzanne Jak, Katharina Meitinger, Natalja Menold, Bengt Muthén, and et al. 2023. Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research* 110: 102805. [CrossRef] [PubMed]
- Little, Roderick J. A., and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. New York: Wiley. [CrossRef]
- Liu, Xinyang, Mattis Geiger, Changsong Zhou, and Andrea Hildebrandt. 2022. Individual differences in white matter microstructure of the face processing brain network are more differentiated from global fibers with increasing ability. *Scientific Reports* 12: 14075. [CrossRef]
- Lodi-Smith, Jennifer, Jonathan D. Rodgers, Valeria Marquez Luna, Sarah Khan, Caleb J. Long, Karl F. Kozlowski, James P. Donnelly, Christopher Lopata, and Marcus L. Thomeer. 2021. The relationship of age with the autism-spectrum quotient scale in a large sample of adults. *Autism in Adulthood* 3: 147–56. [CrossRef]
- Madole, James W., Mijke Rhemtulla, Andrew D. Grotzinger, Elliot M. Tucker-Drob, and K. Paige Harden. 2019. Testing cold and hot cognitive control as moderators of a network of comorbid psychopathology symptoms in adolescence. *Clinical Psychological Science* 7: 701–18. [CrossRef] [PubMed]
- Mellenbergh, Gideon J. 1989. Item bias and item response theory. *International Journal of Educational Research* 13: 127–43. [CrossRef]
- Meredith, William. 1993. Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58: 525–43. [CrossRef]
- Merkle, Edgar C., and Achim Zeileis. 2013. Tests of measurement invariance without subgroups: A generalization of classical methods. *Psychometrika* 78: 59–82. [CrossRef] [PubMed]
- Millsap, Roger E. 2011. *Statistical Approaches to Measurement Invariance*. New York: Routledge. [CrossRef]
- Molenaar, Dylan. 2021. A flexible moderated factor analysis approach to test for measurement invariance across a continuous variable. *Psychological Methods* 26: 660–79. [CrossRef]
- Molenaar, Dylan, and Conor V. Dolan. 2012. Substantively motivated extensions of the traditional latent trait model. *Netherlands Journal of Psychology* 67: 48–57. Available online: <https://bit.ly/3Tu30oV> (accessed on 3 June 2023).
- Molenaar, Dylan, Conor V. Dolan, and Han L. J. van der Maas. 2011. Modeling ability differentiation in the second-order factor model. *Structural Equation Modeling: A Multidisciplinary Journal* 18: 578–94. [CrossRef]
- Molenaar, Dylan, Conor V. Dolan, and Norman D. Verhelst. 2010a. Testing and modelling non-normality within the one-factor model. *British Journal of Mathematical and Statistical Psychology* 63: 293–317. [CrossRef] [PubMed]
- Molenaar, Dylan, Conor V. Dolan, Jelte M. Wicherts, and Han L. J. van der Maas. 2010b. Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence* 38: 611–24. [CrossRef]



- Molenaar, Dylan, Natasa Kő, Sandor Rőzsa, and Andrea Mészáros. 2017. Differentiation of cognitive abilities in the WAIS-IV at the item level. *Intelligence* 65: 48–59. [CrossRef]
- Nowok, Beata, Gillian M. Raab, and Chris Dibben. 2016. synthpop: Bespoke creation of synthetic data in R. *Journal of Statistical Software* 74: 1–26. [CrossRef]
- Oberski, Daniel L. 2013. Individual differences in structural equation model parameters. *arXiv* arXiv:1304.3608.
- Olaru, Gabriel, and Mathias Allemand. 2022. Correlated personality change across time and age. *European Journal of Personality* 36: 729–49. [CrossRef]
- Olaru, Gabriel, and Kristin Jankowsky. 2022. The HEX-ACO-18: Developing an age-invariant HEXACO short scale using ant colony optimization. *Journal of Personality Assessment* 104: 435–46. [CrossRef]
- Olaru, Gabriel, Alexander Robitzsch, Andrea Hildebrandt, and Ulrich Schroeders. 2020. Local structural equation modeling for longitudinal data. *PsyArXiv*. April 24. [CrossRef]
- Olaru, Gabriel, Alexander Robitzsch, Andrea Hildebrandt, and Ulrich Schroeders. 2022. Examining moderators of vocabulary acquisition from kindergarten through elementary school using local structural equation modeling. *Learning and Individual Differences* 95: 102136. [CrossRef]
- Olaru, Gabriel, Ulrich Schroeders, Johanna Hartung, and Oliver Wilhelm. 2019. Ant colony optimization and local weighted structural equation modeling. A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality* 33: 400–19. [CrossRef]
- Olaru, Gabriel, Ulrich Schroeders, Oliver Wilhelm, and Fritz Ostendorf. 2019. ‘Grandpa, do you like roller coasters?’: Identifying age-appropriate personality indicators. *European Journal of Personality* 33: 264–78. [CrossRef]
- Panayiotou, Margarita, Johanna C. Badcock, Michelle H. Lim, Michael J. Banissy, and Pamela Qualter. 2022. Measuring loneliness in different age groups: The measurement invariance of the UCLA loneliness scale. *Assessment*, Epub ahead of print. [CrossRef]
- Putnick, Diane L., and Marc H Bornstein. 2016. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review* 41: 71–90. [CrossRef]
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team. Available online: <https://www.R-project.org/> (accessed on 15 March 2023).
- Reiter, Jerome P. 2023. Synthetic data: A look back and a look forward. *Transactions on Data Privacy* 16: 15–24. Available online: <http://www.tdp.cat/issues21/tdp.a457a22.pdf> (accessed on 3 June 2023).
- Robitzsch, Alexander. 2023a. Model-robust estimation of multiple-group structural equation models. *Algorithms* 16: 210. [CrossRef]
- Robitzsch, Alexander. 2023b. Sirt: Supplementary Item Response theory Models. R Package Version 3.13-228. Available online: <https://CRAN.R-project.org/package=sirt> (accessed on 11 August 2023).
- Robitzsch, Alexander, and Oliver Lüdtke. 2023. Why full, partial, or approximate measurement invariance are not a prerequisite for meaningful and valid group comparisons. *Structural Equation Modeling: A Multidisciplinary Journal*, Epub ahead of print. [CrossRef]
- Rosseel, Yves. 2012. lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48: 1–36. [CrossRef]
- Schroeders, Ulrich, and Malte Jansen. 2022. Science self-concept—more than the sum of its parts? *The Journal of Experimental Education* 90: 435–51. [CrossRef]
- Schroeders, Ulrich, Stefan Schipolowski, and Oliver Wilhelm. 2015. Age-related changes in the mean and covariance structure of fluid and crystallized intelligence in childhood and adolescence. *Intelligence* 48: 15–29. [CrossRef]
- Seifert, Ingo S., Julia M. Rohrer, Boris Egloff, and Stefan C. Schmukle. 2022. The development of the rank-order stability of the big five across the life span. *Journal of Personality and Social Psychology* 122: 920–41. [CrossRef] [PubMed]
- Shapiro, Alexander. 2012. Statistical inference of covariance structures. In *Current Topics in the Theory and Application of Latent Variable Models*. Edited by M. C. Edwards and R. C. MacCallum. London: Routledge, pp. 222–40. [CrossRef]
- Silverman, Bernard W. 1986. *Density Estimation for Statistics and Data Analysis*. Boca Raton: CRC Press. [CrossRef]
- Stefanski, Leonard A., and Dennis D. Boos. 2002. The calculus of M-estimation. *The American Statistician* 56: 29–38. [CrossRef]
- Tucker-Drob, Elliot M. 2009. Differentiation of cognitive abilities across the life span. *Developmental Psychology* 45: 1097–118. [CrossRef] [PubMed]
- Van den Akker, Alithe L., Daniel A. Briley, Andrew D. Grotzinger, Jennifer L. Tackett, Elliot M. Tucker-Drob, and K. Paige Harden. 2021. Adolescent big five personality and pubertal development: Pubertal hormone concentrations and self-reported pubertal status. *Developmental Psychology* 57: 60–72. [CrossRef]
- Wagner, Jenny, Oliver Lüdtke, and Alexander Robitzsch. 2019. Does personality become more stable with age? Disentangling state and trait effects for the big five across the life span using local structural equation modeling. *Journal of Personality and Social Psychology* 116: 666–80. [CrossRef]
- Wang, Ting, Edgar C. Merkle, and Achim Zeileis. 2014. Score-based tests of measurement invariance: Use in practice. *Frontiers in Psychology* 5: 438. [CrossRef] [PubMed]
- Watrin, Luc, Ulrich Schroeders, and Oliver Wilhelm. 2022. Structural invariance of declarative knowledge across the adult lifespan. *Psychology and Aging* 37: 283–97. [CrossRef]
- Weiss, Selina, Diana Steger, Ulrich Schroeders, and Oliver Wilhelm. 2020. A reappraisal of the threshold hypothesis of creativity and intelligence. *Journal of Intelligence* 8: 38. [CrossRef]
- White, Halbert. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50: 1–25. [CrossRef]

- Whitley, Elise, Ian J. Deary, Stuart J. Ritchie, G. David Batty, Meena Kumari, and Michaela Benzeval. 2016. Variations in cognitive abilities across the life course: Cross-sectional evidence from understanding society: The UK household longitudinal study. *Intelligence* 59: 39–50. [[CrossRef](#)] [[PubMed](#)]
- Yuan, Ke-Hai, and Peter M. Bentler. 2007. Structural equation modeling. In *Handbook of Statistics, Vol. 26: Psychometrics*. Edited by C. Randhakrishna Rao and Sandip Sinharay. pp. 297–358. [[CrossRef](#)]
- Zheng, Anqing, Daniel A. Briley, Margherita Malanchini, Jennifer L. Tackett, K. Paige Harden, and Elliot M. Tucker-Drob. 2019. Genetic and environmental influences on achievement goal orientations shift with age. *European Journal of Personality* 33: 317–36. [[CrossRef](#)] [[PubMed](#)]

**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.