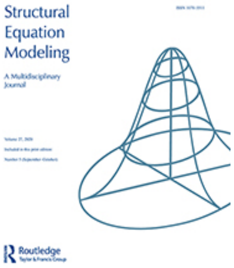

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

Multilevel Modeling in the 'Wide Format' Approach with Discrete Data: A Solution for Small Cluster Sizes

M.T. Barendse & Y. Rosseel

Structural Equation Modeling: A Multidisciplinary Journal, 27: 696–721, 2020

ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2019.1689366

Multilevel Modeling in the 'Wide Format' Approach with Discrete Data: A Solution for Small Cluster Sizes

M.T. Barendse and Y. Rosseel
Ghent University

In multilevel data, units at level 1 are nested in clusters at level 2, which in turn may be nested in even larger clusters at level 3, and so on. For continuous data, several authors have shown how to model multilevel data in a 'wide' or 'multivariate' format approach. We provide a general framework to analyze random intercept multilevel SEM in the 'wide format' (WF) and extend this approach for discrete data. In a simulation study, we vary response scale (binary, four response options), covariate presence (no, between-level, within-level), design (balanced, unbalanced), model misspecification (present, not present), and the number of clusters (small, large) to determine accuracy and efficiency of the estimated model parameters. With a small number of observations in a cluster, results indicate that the WF approach is a preferable approach to estimate multilevel data with discrete response options.

Keywords: Discrete data, multilevel, structural equation modeling, random intercepts

INTRODUCTION

Structural equation modeling (SEM) is a flexible and powerful framework to analyze the interrelationships between many observed and latent variables (see Bollen, 1989). The popularity of SEM is a striking feature of quantitative research in the social sciences (see Guo, Perron, & Gillespie, 2008; MacCallum & Austin, 2000; Yang, 2018). Throughout the years, the SEM framework has been extended to analyze discrete response options, such as binary- or four-point response scales (see Jöreskog & Moustaki, 2001; Wirth & Edwards, 2007) or data from a hierarchical or multilevel structure, such as employees in teams, patients from different doctors, or students in schools (see Hox, Moerbeek, & Van de Schoot, 2017). However, real data are even more complex and often contain a combination of multilevel structures and discrete response options. This calls for more sophisticated estimation techniques.

Three estimation methods have been suggested in the literature for analyzing multilevel SEM data with discrete response options: (1) marginal maximum likelihood (MML; see Hedeker & Gibbons, 1994), (2) Bayesian estimation (e.g., Fox, 2010), and, (3) the multilevel (weighted) least squares estimation method (Asparouhov & Muthén, 2007). Unfortunately, both MML and Bayesian estimation are computationally very intensive, limiting their practical use (see Fox, 2010; Jöreskog & Moustaki, 2001; Wirth & Edwards, 2007). The multilevel (weighted) least squares method can handle many more latent variables, but has not been thoroughly studied and showed mixed results in simulation studies (Asparouhov & Muthén, 2007; Depaoli & Clifton, 2015; Holtmann, Koch, Lochner, & Eid, 2016).

As analyzing multilevel data with discrete response options is complex, researchers often turn to suboptimal analyzes techniques, such as using sum-scores or ignoring either the multilevel or discrete nature of the data (see for example Koomen, Verschueren, van Schooten, Jak, & Pianta, 2012; Lee, 2009; Li, Fortner, & Lei, 2015). This is unfortunate as this may jeopardize the results of the analysis and therefore any decision-making based on these results. Currently lacking are statistical techniques that offer an efficient and practical method to analyze multilevel datasets with discrete response scales.

In this article, we propose a solution for discrete multilevel data with a small number of observations in a cluster (approximately < 10). The National Longitudinal Study of

Correspondence should be addressed to M. T. Barendse. Email: mariska.barendse@ugent.be

Supplemental data for this article can be accessed here.

Adolescent Health (NLSAH, 2005) data is an example of a dataset with small cluster sizes, where at each wave 15,000 adolescents are nested in approximately 2,600 neighborhoods (i.e, about six per cluster). On a smaller scale, among others, Koomen, Verschueren, and Pianta (2007) collected data in Dutch primary schools with on average two to three children in each class. To analyze these type of data, we will use the so called 'wide format' (WF) approach. Using different model specifications, the WF approach is discussed in the SEM literature for continuous multilevel data by Bauer (2003), Curran (2003), and Mehta and Neale (2005). In the WF approach, data are arranged in such a way that each data row corresponds to a single cluster. The multilevel structure is then explicitly modeled as will be explained in section "Multilevel SEM in the WF Approach" and section "WF Approach with Discrete Data." The WF approach differs from the general way multilevel data are organized and modeled, in this article referred to as the 'long format' (LF) approach. In this LF approach each row corresponds to the data of a single unit, and many rows constitute a single cluster. Dealing with discrete data using the WF approach yields some additional advantages over the LF approach. First, the WF can handle multilevel data using single-level software. Second, the multilevel model is explicitly modeled and therefore offers freedom to freely estimate parameters that are restricted in the LF approach.

The goal of this paper is twofold: (1) to offer a general frequentist framework to analyze random intercept multilevel data in the WF approach and (2) to extent and test the WF approach for discrete data. The former goal is necessary as different aspects of the WF approach, like the inclusion of covariates, obtaining test statistics, dealing with unbalanced data, and handling missing data have been described separately in the literature; either in the context of generalized linear mixed models (e.g., Croon & van Veldhoven, 2007), or in the context of WF multilevel models with different specifications (e.g., Bauer, 2003; Curran, 2003; Mehta & Neale, 2005). This fragmented literature has not been united to a general WF framework. To the best of our knowledge, an extension of the WF approach to discrete data has not been investigated yet. As will be explained in section "WF Approach with Discrete Data" and the illustrative example, the WF approach turns out to be much more computationally efficient compared to the often used MML estimation method in the LF approach. Using the WF approach does not require specialized multilevel software, which decreases the complexity of estimating multilevel models.

The paper is organized as follows: First, a brief overview of single level SEM and multilevel SEM for continuous and discrete data is given. Next, we will describe the WF approach for continuous and discrete data. Thereafter, we will perform a simulation study to evaluate the general WF framework and to compare several WF estimation methods to LF estimation methods in terms of the accuracy and efficiency of the parameter estimates under different conditions. Finally, the use of the WF approach for discrete data will be illustrated by means of an application in educational research on student-teacher relationships.

Single-level SEM

Continuous Data

The measurement equation and the structural equation are the two fundamental equations that define the general SEM framework. The measurement equation equals

$$\mathbf{y} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\eta + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is a $p \times 1$ vector of observed variables, $\boldsymbol{\nu}$ is a $p \times 1$ vector of intercepts, $\boldsymbol{\Lambda}$ is a $p \times m$ matrix of factor loadings relating the p observed variables to the m latent variables, η is a $m \times 1$ vector of latent variable scores, and $\boldsymbol{\varepsilon}$ is a $p \times 1$ vector of residuals. The structural equation equals

$$\eta = \boldsymbol{\alpha} + \mathbf{B}\eta + \boldsymbol{\zeta}, \tag{2}$$

where $\boldsymbol{\alpha}$ is a $m \times 1$ vector of latent factor means and intercepts, $\mathbf{B}$ is a $m \times m$ matrix of regression coefficients among the latent factors, and $\boldsymbol{\zeta}$ is a $m \times 1$ vector of residuals. In SEM we assume that $\mathrm{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\zeta}) = \mathbf{0}$, $\mathrm{Cov}(\eta, \boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{Cov}(\eta, \boldsymbol{\zeta}) = \mathbf{0}$, $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{E}(\boldsymbol{\zeta}) = \mathbf{0}$, $\mathrm{diag}(\mathbf{B}) = \mathbf{0}$, and that $(\mathbf{I} - \mathbf{B})$ is invertible, where $\mathbf{I}$ is a $m \times m$ identity matrix. Using covariance algebra we can find expressions for the covariance and the mean structure of $\mathbf{y}$, denoted by $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$. The $p \times p$ covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{y}$ is expressed as a function of the model parameters ($\boldsymbol{\theta}$) and equals

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})^{-1T}\boldsymbol{\Lambda}^T + \boldsymbol{\Theta} \tag{3}$$

where variances and covariances of η and $\boldsymbol{\varepsilon}$ are denoted by $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$. The mean structure implied by Equation 1 and Equation 2 equals

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha}. \tag{4}$$

Assuming multivariate normality, the model parameters ($\boldsymbol{\theta}$) can be estimated via maximizing the likelihood, or equivalently, by minimizing the following objective function:

$$\begin{aligned} F_{ML} = &\{ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - ln|\mathbf{S}| + tr[\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{S}] - p\} \\ &+ \{[\bar{\mathbf{y}} - \boldsymbol{\mu}(\boldsymbol{\theta})]^T\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}[\bar{\mathbf{y}} - \boldsymbol{\mu}(\boldsymbol{\theta})]\}, \end{aligned} \tag{5}$$

where $\mathbf{S}$ is the sample covariance matrix and $\bar{\mathbf{y}}$ is the observed mean vector. Under the assumptions of multivariate normality, a large enough sample size and a correct model specification yields unbiased parameter estimates (Bollen, 1989). Alternative estimators are the generalized least squares and the weighted least squares (see Bollen, 1989; Browne, 1995; Jöreskog, 1981).

Discrete Data

Items are often scored with binary codings or three- or four-point scales. To deal with these discrete data, a distinction is made in the SEM literature between limited information estimation methods that use only summaries of the data and full information estimation methods that use all available information in the data.

Limited Information Estimation Methods. The least squares estimation methods and the pairwise maximum likelihood (PML) estimation method are both limited information methods that use only summaries of the data. These methods, mostly developed within the SEM literature, assume underlying continuous latent response variables. A variable y_{ik} for a certain individual on item k with C_k response scales stems from an underlying continuous variable y^*_{ik} with a normal distribution $N(y^*_{ik}|0, \sigma^2_k)$ and τ_{kc} values that refer to thresholds

$$y_{ik} = c \Leftrightarrow \tau_{k,c-1} < y^*_{ik} < \tau_{k,c} \tag{6}$$

for categories $c_k = 1, 2, \ldots, C_k$, with $\tau_{k,0} = -\infty$ and $\tau_{k,C} = +\infty$.

Least Squares Estimation Methods. The three stage weighted least squares method is often applied to estimate models for large datasets with discrete data. In the first stage, the thresholds are estimated using the univariate data. In the second stage, the $p(p-1)/2$ polychoric correlations are estimated (see Olsson, 1979; Olsson, Drasgow, & Dorans, 1982). In the third stage, the model parameters are estimated using the weighted least squares estimation method;

$$F_{WLS} = (\mathbf{s} - \hat{\boldsymbol{\sigma}})' \mathbf{W}^{-1} (\mathbf{s} - \hat{\boldsymbol{\sigma}}), \tag{7}$$

where $\mathbf{s}$ denotes a vector with non-redundant sample-based statistics, $\hat{\sigma}$ denotes a vector with the non-redundant model-based statistics (i.e., thresholds and polychoric correlations), and $\mathbf{W}^{-1}$ denotes the inverse of a weight matrix that estimates the asymptotic covariance matrix of $\sqrt{I}\mathbf{s}$ (see Muthén & Satorra, 1995), where I denotes the sample size. In the least squares framework (see Browne, 1984; Muthén, Du Toit, & Spisic, 1997), there are three different choices for $\mathbf{W}$, leading to WLS using the full weight matrix $\mathbf{W}$, DWLS using only the diagonal of $\mathbf{W}$, and ULS where $\mathbf{W}$ is replaced by the identity matrix ($\mathbf{I}$).

Pairwise Maximum Likelihood Estimation Method. The single level PML estimation method was introduced in the SEM framework by Jöreskog and Moustaki (2001). In this estimation method, the complex likelihood is broken down as a product of bivariate (and sometimes univariate) likelihoods which are computationally easier to handle. The PML estimation method is part of a broader framework of composite ML estimators which are asymptotically unbiased, consistent, and normally distributed (see Lindsay, 1988; Varin, 2008). In PML, the log-likelihood contribution of a single observation is a sum of $p^\star = p(p-1)/2$ components, each component being the bivariate log-likelihood of two variables (i.e., k and l):

$$\begin{aligned} \log\ l_i &= \sum_{k=1}^{p-1} \sum_{l=k+1}^{p} [\log f(y_{ik}, y_{il}; \boldsymbol{\theta})] \\ &= \sum_{k<1} [\log\ f(y_{ik},\ y_{il};\ \boldsymbol{\theta})]. \end{aligned} \tag{8}$$

Given a random sample $\mathbf{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots \boldsymbol{y}_I\}$ of size I, the total log-likelihood of the data is the sum of all individual contributions:

$$\log\ L(\boldsymbol{\theta}; \mathbf{Y}) = \sum_{i=1}^{I} \log l_i. \tag{9}$$

In principle Equation 8 is general and can deal with any type of data (continuous or discrete, and combinations thereof). Until now, the PML estimation method in the SEM context has only been applied with discrete data. For discrete indicators k and l, the exact form of $f(y_{ik}, y_{il}; \boldsymbol{\theta})$ in Equation 8 is:

$$\begin{aligned} \log\ f(y_{ik}, y_{il}; \boldsymbol{\theta}) = \textstyle\sum_{a=1}^{C_k} \sum_{b=1}^{C_l} I(y_{ik} = a, y_{il} = b) \\ \log \omega(y_{ik} = a, y_{il} = b); \boldsymbol{\theta}), \end{aligned} \tag{10}$$

where

$$\begin{aligned} \omega(y_{ik} = a, y_{il} = b; \boldsymbol{\theta}) &= \int_{\tau_{k,a-1}}^{\tau_{k,a}} \int_{\tau_{l,b-1}}^{\tau_{l,b}} f(y^*_{ik}, y^*_{il}; \boldsymbol{\theta}) dy^*_{ik} dy^*_{il}, \\ &= \phi_2(\tau_{k,a}, \tau_{l,b}; \rho_{kl}) \\ &\quad - \phi_2(\tau_{k,a-1}, \tau_{l,b}; \rho_{kl}) \\ &\quad - \phi_2(\tau_{k,a}, \tau_{l,b-1}; \rho_{kl}) \\ &\quad + \phi_2(\tau_{k,a-1}, \tau_{l,b-1}; \rho_{kl}), \end{aligned} \tag{11}$$

where ρ_{kl} denotes the model implied correlation between y^*_{ik} and y^*_{il}, and $\phi_2(\tau_1, \tau_2; \rho)$ denotes the bivariate cummulative normal distribution with correlation ρ evaluated at point (τ_1, τ_2). The PML estimator produces unbiased results (see Katsikatsou, Moustaki, Yang-Wallentin, & Jöreskog, 2012).

Scaling with Discrete Data. With discrete data, we need to determine a metric for $\mathbf{y}^*$. A common way is to use the delta parameterization where Θ is not a free parameter, but given by

$$\boldsymbol{\Theta} = \boldsymbol{\Delta}^{-2} - \mathrm{diag}(\boldsymbol{\Sigma}^*) \tag{12}$$

where $\boldsymbol{\Sigma}^*$ equals $\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\mathbf{B})^{-1T}\boldsymbol{\Lambda}^T$ with the scaling factors defined as

$$\boldsymbol{\Delta} = \mathrm{diag}(\boldsymbol{\Sigma}^*)^{-1/2}. \tag{13}$$

Alternatively, the theta parameterization can be chosen. Here the diagonal of $\boldsymbol{\Theta}$ is an identity matrix and the scaling factors ($\boldsymbol{\Delta}$) are obtained as:

$$\boldsymbol{\Delta}^{-2} = \mathrm{diag}(\boldsymbol{\Sigma}^*) + \boldsymbol{\Theta}. \tag{14}$$

Full Information Estimation Method. An often used full information estimation method is the marginal maximum likelihood (MML), as described by Bock and Aitkin (1981). This estimation method has been developed within the item response theory literature and integrates over the distribution of the latent variables. The log-likelihood for the data (i.e., $\mathbf{Y}$) can be written as the sum of all of the log-likelihoods for all individuals i:

$$\log L(\boldsymbol{\theta};\mathbf{Y}) = \sum_{i=1}^{I}\log l_i$$

$$= \sum_{i=1}^{I}\log f_i(\mathbf{y}_i;\boldsymbol{\theta}_y,\boldsymbol{\theta}_\eta) \tag{15}$$

where the individual likelihood contribution of observation i equals

$$l_i = \int_{D(\eta)} f(\mathbf{y}_i|\boldsymbol{\eta};\boldsymbol{\theta}_y)g(\boldsymbol{\eta};\boldsymbol{\theta}_\eta)d\boldsymbol{\eta}$$

$$= \int_{D(\eta)}\prod_{k=1}^{p} f_{ik}(y_{ik}|\boldsymbol{\eta};\boldsymbol{\theta}_y)g(\boldsymbol{\eta};\boldsymbol{\theta}_\eta)d\boldsymbol{\eta} \tag{16}$$

and where $D(\eta)$ is the domain of integration and $g(\bullet)$ is the prior density of η. The MML estimation method is computationally very intensive as it has to integrate out the latent variables using numerical integration (i.e., Gauss-Hermite quadrature, adaptive quadrature, Laplace approximation, or Monte Carlo integration). This estimation method can be applied with different link functions (e.g., probit, logit, log, logistic, and complementary log-log).

Multilevel SEM in the LF Approach

From the late 1980’s, the SEM framework started to incorporate features from multilevel regression (e.g., Bryk & Raudenbush, 1987). Schmidt (1969) was the first to describe a saturated model of a within-covariance matrix (i.e., individual deviations from the group mean) and a between-covariance matrix (i.e., group means) for balanced data. His work was further developed by Goldstein and McDonald (1988), McDonald and Goldstein (1989), Muthén (1989, 1990), and McDonald (1993). More recently, estimation methods to analyze multilevel discrete data were developed. Multilevel SEM in the LF approach is the standard way to model multilevel data where the data is organized in such a way that each row corresponds to a single unit.

Continuous Data

Here, we will describe the multilevel model as introduced by McDonald and Goldstein (1989) with a slightly adjusted mean structure. In a two-level set-up, the multivariate response vector $\mathbf{y}_{ji}$ can be decomposed in a between part $\mathbf{u}_j$ and a within part $\mathbf{u}_{ji}$

$$\mathbf{y}_{ji} = \mathbf{u}_j + \mathbf{u}_{ji}, \tag{17}$$

where $j = 1,\ldots,J$ is an index for the clusters, and $i = 1,\ldots,I_j$ is an index for the units within a cluster. As $\mathbf{u}_j$ and $\mathbf{u}_{ji}$ are independent, the expected value equals

$$\mathrm{E}(\mathbf{y}_{ji}) = \boldsymbol{\mu}_b + \boldsymbol{\mu}_w, \tag{18}$$

where $\boldsymbol{\mu}_w$ is only needed if there are variables that only exist at the within-level. The covariance can be decomposed as

$$\mathrm{Cov}(\mathbf{y}) = \boldsymbol{\Sigma}_T = \boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_w. \tag{19}$$

On both the within-level ($\boldsymbol{\Sigma}_w$, $\boldsymbol{\mu}_w$) and the between-level ($\boldsymbol{\Sigma}_b, \boldsymbol{\mu}_b$) a different model can be fitted. Consider data from a single cluster with $\mathbf{z}$ representing variables at the between-level:

$$\mathbf{v}_j = [\mathbf{z}_j, \mathbf{y}_{j1}, \mathbf{y}_{j2}\ldots\mathbf{y}_{jI_j}]^T, \tag{20}$$

with the following expectation of $\mathbf{v}_j$

$$\mathrm{E}[\mathbf{v}_j] = \bar{\mathbf{v}}_j = [\mathbf{z}_j, \mathbf{y}_j]^T = [\mathbf{z}_j, \mathbf{y}_{j1}, \mathbf{y}_{j2}\ldots\ \mathbf{y}_{jI_j}]^T. \tag{21}$$

The covariance matrix, referred to as $\mathbf{V}_j(\boldsymbol{\theta})$, can be written as:

$$Cov[\mathbf{v}_j] = \mathbf{V}_j(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{zz} & \mathbf{1}_{I_j}^T \otimes \boldsymbol{\Sigma}_{zy} \\ \mathbf{1}_{I_j} \otimes \boldsymbol{\Sigma}_{yz} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}. \tag{22}$$

with $\boldsymbol{\Sigma}_{zz} = \mathrm{Cov}(z_j, z_j)$, $\boldsymbol{\Sigma}_{zy} = \mathrm{Cov}(z_j, \mathbf{y}_j)$, and $\boldsymbol{\Sigma}_{yy} = \mathrm{I}_{I_j} \otimes \boldsymbol{\Sigma}_w(\boldsymbol{\theta}) + \mathbf{1}_{I_j}\mathbf{1}_{I_j}^T \otimes \boldsymbol{\Sigma}_b(\boldsymbol{\theta})$. Assuming multivariate normality, the model parameters in $\boldsymbol{\theta}$ can be estimated via minimizing the ML objective function which equals minus two times the log–likelihood function without a constant over all clusters:

$$F_{ML} = \sum_{j=1}^{J} ln|\mathbf{V}_j| + (\mathbf{v}_j - \overline{\mathbf{v}}_j)^T \mathbf{V}_j^{-1} (\mathbf{v}_j - \overline{\mathbf{v}}_j). \tag{23}$$

As $\mathbf{V}_j$ can become quite large, several methods are used to find expressions to compute $\mathbf{V}_j^{-1}$ and $|\mathbf{V}_j|$ efficiently (see McDonald & Goldstein, 1989).

Discrete Data

Some estimation methods for single level discrete data can be applied in the multilevel context. As multilevel data has a more complex structure, adjustments to the algorithms have to be made.

Limited Information Estimation Methods. Due to the work of Asparouhov and Muthén (2007) the least squares estimation methods can be applied to discrete multilevel data. For the discrete variable k for individual i in cluster j, the threshold parameters are now defined by

$$y_{ijk} = c \Leftrightarrow \tau_{k,c-1} < y^*_{ijk} < \tau_{k,c} \tag{24}$$

The underlying variable $\mathbf{y}^*_{\mathrm{ij}}$ is decomposed into a between-level ($\mathbf{u}_{\mathrm{bj}}$; i.e., random intercept) and a within-level ($\mathbf{u}_{\mathrm{wij}}$) part:

$$\boldsymbol{y}^*_{ij} = \boldsymbol{u}_{wij} + \boldsymbol{u}_{bj}. \tag{25}$$

A separate structural model can now be defined at both levels. The least squares estimation methods for multilevel data in the LF approach are based on the single level least squares estimation methods. In the first stage, the thresholds are estimated using an univariate model with the two-level maximum likelihood estimation method (Asparouhov & Muthén, 2007). In the second stage, polychoric correlations are estimated based on bivariate models by fixing the univariate parameters to their first stage estimates (Asparouhov & Muthén, 2007). Minimizing the fit function with respect to the parameters of the model at both levels with Equation 7, is the last stage of the estimation process.

Full Information Estimation Method. The LF MML, as described by Hedeker and Gibbons (1994) and Rabe-Hesketh, Skrondal, and Pickles (2004a), is based on the MML for single level data (Bock & Aitkin, 1981). In the multilevel context, the random intercept can be written as a vector of latent variables, referred to as η_y, so that the extended set of latent variables equals $\eta^* = (\eta, \eta_y)$. This formula can be extended to models with more latent variables at different levels of the multilevel model. For a two level model, the log-likelihood for the data can be written as the sum of the log-likelihoods of all the clusters j:

$$\begin{aligned} \log L(\boldsymbol{\theta}; \mathbf{Y}) &= \sum_{j=1}^{J} \log l_j \\ &= \sum_{j=1}^{J} \log f_j(\mathbf{y}_j; \boldsymbol{\theta}_y, \boldsymbol{\theta}_{\eta^*}) \end{aligned} \tag{26}$$

where the individual likelihood contribution of cluster j equals

$$\begin{aligned} l_j &= \int_{D(\eta^*)} f(\mathbf{y}_j | \boldsymbol{\eta}^*; \boldsymbol{\theta}_y) g(\boldsymbol{\eta}^*; \boldsymbol{\theta}_{\eta^*}) d\boldsymbol{\eta}^* \\ &= \int_{D(\eta^*)} \prod_{k=1}^{p} f_{jk}(y_{jk} | \boldsymbol{\eta}^*; \boldsymbol{\theta}_y) g(\boldsymbol{\eta}^*; \boldsymbol{\theta}_{\eta^*}) d\boldsymbol{\eta}^* \end{aligned} \tag{27}$$

where $D(\eta^*)$ is the domain of integration and $g(\bullet)$ is the prior density of η^*. The multilevel MML estimation method is very flexible and can be extended with more levels, covariates, different link functions (e.g., probit, logit, log, logistic, and complementary log-log), and random slopes (see Rabe-Hesketh et al., 2004a). Unfortunately, the MML estimation technique for multilevel data is computationally even more intensive than the single level MML estimation method. To reduce the number of latent variables in two-level factor models, the residuals at the between-level are usually not estimated by default (e.g., Muthén & Muthén, 2010; Rabe-Hesketh, Skrondal, & Pickles, 2004b).

Multilevel SEM in the WF Approach

As an alternative to the above described LF approach to fit multilevel data, one can use a WF approach. In this WF approach, the multilevel model is explicitly modeled and the data is restructured such that all units in a cluster are stored in a single row. Despite several papers describing the WF approach, this approach is seldom applied. Below we offer a general framework to fit multilevel data with a random intercept in the WF approach for both continuous and discrete data.

Continuous Data

McArdle and Epstein (1987) were the first authors who showed how multilevel models of individual change can be fitted using a mean and covariance structure analysis with time-structured data in latent growth curve models (see also Chou, Bentler, & Pentz, 1998; MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997; Meredith & Tisak, 1990; Willett & Sayer, 1994). The latent growth curve models become confirmatory factor analysis (CFA) models with model parameters fixed to specific data values. A random-intercept multilevel model can then be estimated as a restricted CFA model, where univariate multilevel models become multivariate uni–level models. In this way, each data row consists of all observations in a cluster with latent variables constructed from unit-specific versions of the variables. Figure 1 shows data, a graphical representation, and a formula of a random intercept model in both the LF approach and the WF approach.

Later, Curran (2003) and Bauer (2003) further developed the WF approach and provided examples of multilevel SEM using the WF approach. Curran extended the growth curve model with more SEM related options, such as random intercepts and slopes, and latent or observed covariates at the between-level. Bauer estimated random intercept models including level-2 predictors and structural relations between latent variables. He also reflected on the use of formal tests of model fit and demonstrated the applicability of multilevel SEM with unbalanced data. In 2005, Mehta and Neale provided a more extended framework for modeling multilevel models in the WF approach with continuous data, where person-specific data is used for modeling means as well as covariances at an individual level. The main difference between the representation of Mehta and Neale and Bauer and Curran is that the models of Mehta and Neale impose restrictions on the covariance matrix in such a way that the model is decomposed into a between–level and a within–level that allows for different structures and parameter values at different levels.

In this paper, we will continue with the model of Mehta and Neale (2005) that explicitly models the multilevel model of McDonald and Goldstein (1989). To deal with continuous data in the WF approach, we proceed with the following steps:

1. Rearrange the data in such a way that each row corresponds to a single cluster (see data in Figure 1).
2. Construct a model at the within-level involving the variables that belong to a single unit in a cluster and repeat this model as many times as the maximum cluster size.
3. Put equality constraints on all parameters across units in a cluster. For example, in a one factor model, equality constraints are necessary on the factor loadings, factor variances, and error variances. If variables are both at the within- and the between-level,

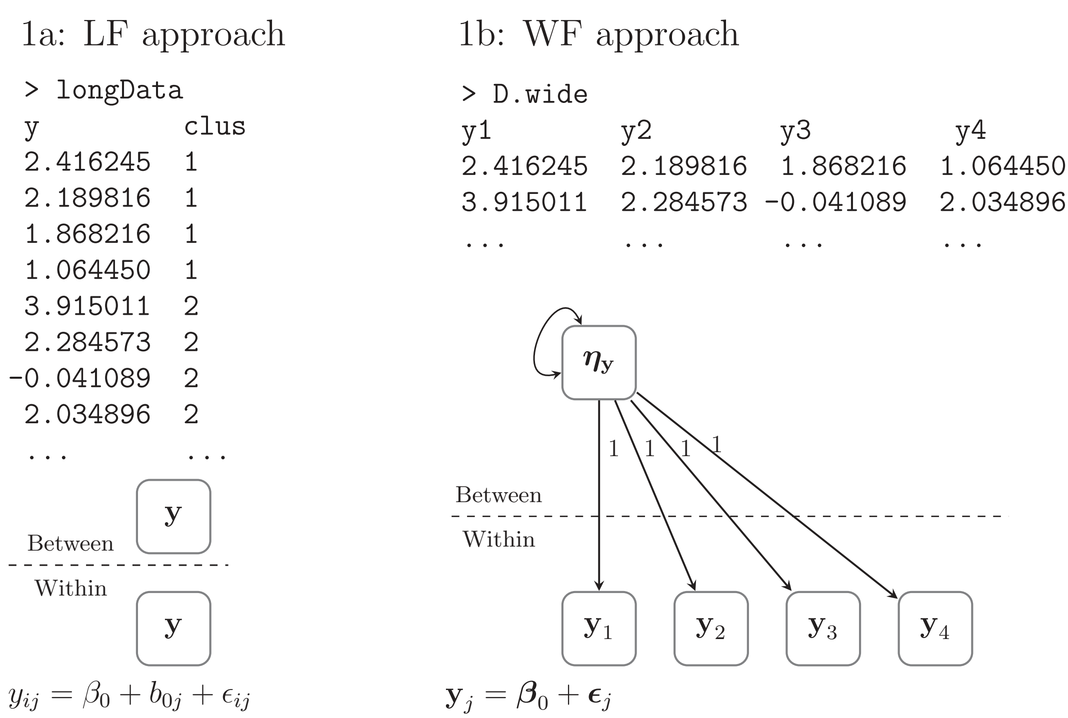

FIGURE 1 A random intercept model with the corresponding data, figure and formula in both the LF (left hand side)- and WF -approach (right hand side). Note: The rounded boxes represent the within- and between versions of the original variable y for four units per cluster. η_y corresponds to the unit-specific version of y on the between level, which equals a factor with loadings fixed at unity. Residual variances (not shown here) contain equality restrictions to ensure only one residual variance is estimated.

the intercepts at the within-level should be fixed to zero.

4. For each endogenous variable at the within-level, construct a latent variable where the indicators correspond to the unit-specific version of that variable; The factor loadings are fixed to one; These latent variables represent the random intercepts of these variables in the model.
5. Construct a model at the between-level with the newly constructed between-level latent variables.

Figure 2 shows a one factor model according to the parameterization of Mehta and Neale (2005) for the case where we have three units in each cluster. The within factor is referred to as η_{fw}, the unit-specific version of the variables on the between-level is referred to as η_y, and the between factor is referred to as η_{fb}. To clarify the proposed steps, we give the lavaan syntax corresponding to Figure 2 in Appendix A. The figure shows that the WF approach is more flexible than the LF approach, as the equality restrictions across units in a cluster (e.g., *e* to *h* for the factor loadings and/or *i* to *l* for the residual variances in Figure 2) can all be tested by freeing the restrictions in the WF approach. Note that within the LF approach, it is not possible to free restrictions across clusters. Below we will describe how to extent the framework of Mehta and Neale (2005) and deal with covariates, missing data, and obtain fit statistics.

WF Approach with a Covariate. The WF multilevel model can be extended with covariates at the within- and/or the between-level (see Croon and van Veldhoven (2007) for adding covariates in a WF multilevel regression model). Values for the between-level covariate are equal for all units in a cluster and can be added to the model. The effect of the between-covariate is calculated by regressing the random intercept on the covariate. Values for the within-level covariate are unique for each individual in a cluster. The effect of the within-covariate is calculated by regressing the observed variables at the within-level on the covariates with equality restrictions. As the within-level covariate contains equality restrictions, they are considered stochastic and jointly modeled with the other variables as endogenous variables.

Lüdtke et al. (2008) describe multilevel models with a covariate that exists both at the within- and the between-

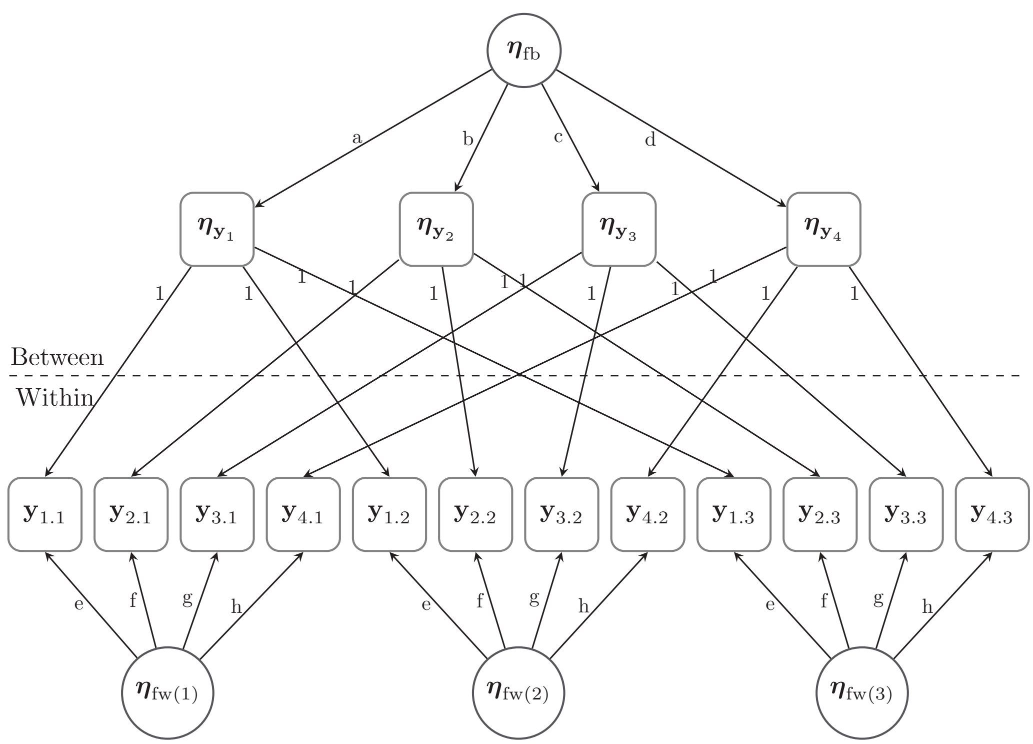

FIGURE 2 One factor multilevel CFA model in the WF approach (four variables with three units per cluster). Note: The rounded boxes represent the within- and between versions of the original variables $\boldsymbol{y}_{1.1}$ to $\boldsymbol{y}_{4.3}$ for three units per cluster. η_{y_1} to η_{y_4} correspond to the unit-specific version of that variable on the between level, where labels *a* to *d* reflect the between-level factor loadings. For identification purposes, one factor loading or the variance of η_{fb} must be fixed to unity. The structures below $\boldsymbol{y}_{1.1}$ to $\boldsymbol{y}_{4.3}$ account for the within–level latent variables (i.e., $\eta_{\text{fw}(1)}$ to $\eta_{\text{fw}(3)}$) with parameter labels *e* to *h* for the factor loadings. For identification purposes, one factor loading of each within factor (η_{fw}) or the variance of η_{fw} must be fixed to unity. Identical parameter labels indicate equality constraints in the model. Residual variances (not shown here) also contain equality restrictions (i.e., *i* to *l*). Appendix A shows the corresponding lavaan syntax.

level in the LF approach. One can then either aggregate the covariate before the analysis to get between-level values for the covariate, referred to as the multilevel manifest covariate approach, or split the covariate in a within and a between part, referred to as the multilevel latent covariate approach. Figure 3a shows the MMC approach in WF approach and Figure 3b shows the MLC approach in the ‘wide format approach. To obtain the correct between-level effect in Figure 3b, one has to calculate the so called “contextual effect” by estimating the between–level effect (i.e., b) and subtract this from the within-level effect (i.e., a). Appendix B shows the lavaan syntax that can be added to Appendix A to fit a model with a covariate on both the within- and the between-level.

Missingness in the WF Approach. In the multilevel SEM setting, we make a distinction between design missingness and within cluster missingness. Design missingness occurs with an unbalanced multilevel design, where the cluster sizes differ. In the WF approach the cluster with the largest number of units determines the width of the dataset and all other clusters with less units have missing values (see Bauer, 2003). With cluster missingness, we have missing values for some variables within a complete cluster. If we assume missing at random (MAR), we can deal with both types of missing data by using full information maximum likelihood (Arbuckle, 1996; Neale, 2000).

Evaluation of Fit and Calculation of Intra Class Correlation. Both the evaluation of fit and the calculation of the intra class correlations (ICC) in the WF approach depend on the estimation of an unstructured or saturated (H1) model. The arbitrary ordering of the units in a multilevel model needs some modification in the unstructured model. The notion of interchangeability, which entails that there is nothing in the model that distinguishes one unit from another within a given cluster, is important here (Bauer, 2003). The appropriate unstructured model is more complex and contains equality restrictions across units in a cluster, following the notion that units are

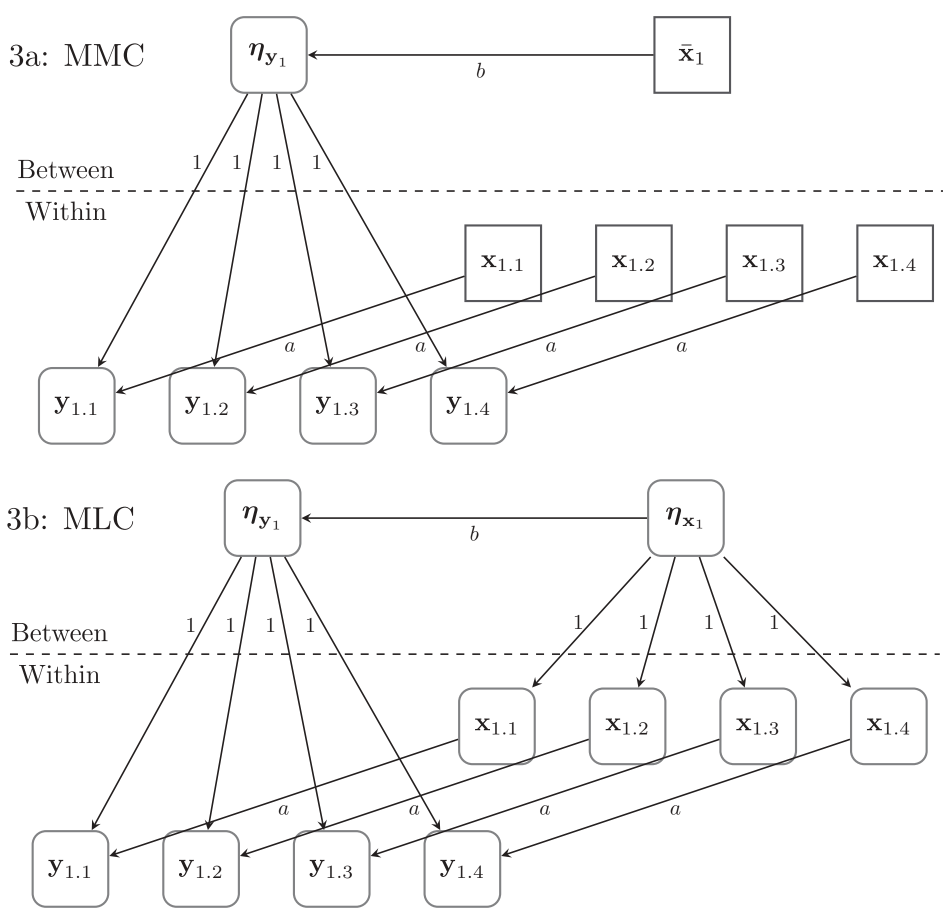

FIGURE 3 A random intercept path model in the WF approach with four items and a within– and a between-covariate modeled in two distinct ways. Note: η_y represents a random intercept and η_x represents an aggregated covariate on the between-level. Identical parameter labels a indicate equality constraints in the model. MMC refers to the multilevel manifest covariate approach (Figure 3a) and MLC refers to multilevel latent covariate approach (Figure 3b). The rounded boxes represent the within- and between versions of the original variables. Residual variances (not shown here) contain equality restrictions.

essentially interchangeable. In the calculation of a fit statistic, the model of interest will then be compared with the unstructured model (H1) that serves as a benchmark of perfect fit that takes the multilevel structure into account. Figure 4 shows, for example, an unrestricted model with three units per cluster with the associated equality restrictions to describe the multilevel structure. This model can be used to calculate the chi-square value of the one-factor model shown in Figure 2. If models also include a covariate, the within-level and/or between-level covariates are also part of the unrestricted model (H1).

The intra class correlation (ICC), that calculates how much variance in a response variable stems from between group differences, also needs the estimation of an unstructured model in the WF approach. For the calculation of the ICC one only needs the variance of each variable at both levels in the unstructured model (ICC = Var between/(Var between + Var within); see Hox et al., 2017).

WF Approach with Discrete Data

Although the discrete nature of the data has to be taken into account, the general procedure to estimate multilevel models with discrete data in the WF approach is somewhat similar to the estimation of continuous data in the WF approach. With continuous responses, the WF approach and the LF approach result in identical parameter estimates and standard errors. However, with discrete data the WF estimates are no longer identical to the LF estimates. We will use the same stepwise procedure as the one described with continuous data. Step 1 to Step 3 and Step 5 are identical, but Step 4 needs to be adjusted with discrete data:

4. For each endogenous variable at the within-level, construct a latent variable where the indicators correspond to the unit-specific version of that variable; The factor loadings are fixed to one; These latent variables represent the random intercepts of these variables in the model. The thresholds are restricted according to a similar structure with unit-specific restrictions.

Modeling the between-level with discrete data is equivalent to modeling the between-level with continuous data, but the within-level needs to be adjusted to take the thresholds into account. Figure 5 represents, for example, a very simple random intercept only model with four response options. Both the theta- and delta parameterization are applicable to fit multilevel data in the WF approach. With discrete response options, equality restrictions on the thresholds across units are necessary. As Figure 5 has four response options, three thresholds are estimated. To illustrate how this works in practice, Appendix C gives the corresponding (stepwise) lavaan syntax in the theta parameterization.

More complex models than the one shown in Figure 5 can be estimated. Appendix D shows how to fit a one factor

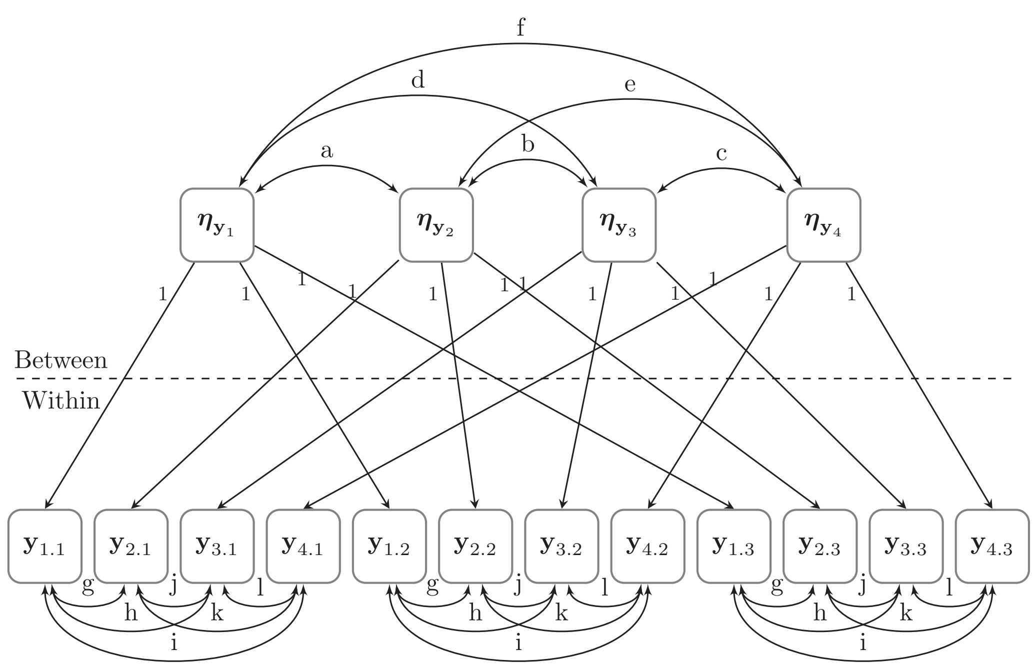

FIGURE 4 Unrestricted model in the WF approach (four variables with three units per cluster). Note: The rounded boxes represent the within- and between versions of the original variables $\boldsymbol{y}_{1.1}$ to $\boldsymbol{y}_{4.3}$ for three units per cluster. η_{y_1} to η_{y_4} correspond to the unit-specific version of that variable on the between level, where labels a to f reflect the between-level covariances. The structures below $\boldsymbol{y}_{1.1}$ to $\boldsymbol{y}_{4.3}$ account for the within–level latent variables (i.e., $\eta_{\mathrm{fw}(1)}$ to $\eta_{\mathrm{fw}(3)}$) with parameter labels g to l reflecting the covariances on the within-level. Identical parameter labels indicate equality constraints in the model. Residual variances are not shown here, but also contain equality restrictions.

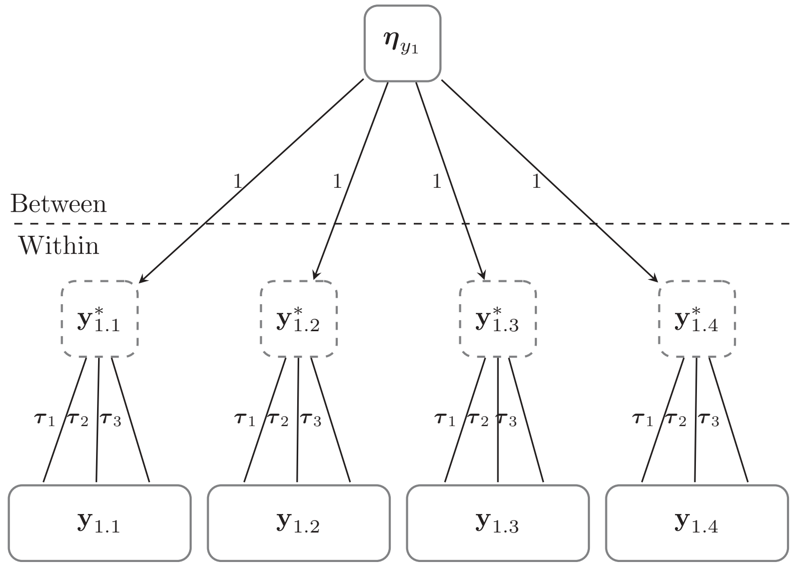

FIGURE 5 Random intercept only model with four items. Note: (η_{y_1}) refers to a random intercept, τ refers to the thresholds, and y^* refers to the underlying latent response variables. Identical parameter labels indicate equality constraints.

model with discrete data in lavaan according to the proposed steps . Similar to the continuous case, covariates can be added to the model (see Appendix B and Figure 3a and b). Adding (random) discrete covariates can be an area of concern in multilevel analysis, as discrete covariates should be considered discrete and not continuous. The WF approach can also add discrete covariates by estimating thresholds for the covariate. This is in contrast with for example multilevel least squares estimation methods in the LF approach, where discrete covariates are considered continuous. In the LF MML estimation approach the covariates are usually considered as fixed, and are regressed out, rendering the distinction between discrete (dummy) and continuous covariates irrelevant.

In theory, all single level estimators for discrete data can be used to estimate multilevel models in the WF approach. However, model estimation, missing data procedures, and the calculation of fit statistics are dependent on the chosen estimation method, as briefly outlined below.

Least Squares Estimation Methods

Model Estimation. The least squares estimation methods (i.e., ULS, WLS, DWLS) are able to estimate all multilevel models in the WF approach.

Missing Data. As the least squares estimation method is a multiple step estimation method, the procedure for missing data also consists of different steps. In the first step, the univariate information is analyzed and listwise deletion is applied. In the second step, only the cases where both variables are observed contribute to the calculation of the bivariate likelihood for estimation of the polychoric and/or polyserial correlations. This is referred to as the complete-pairs (CP) procedure.

Evaluation of Fit and ICC. Similar to the calculation of a fit statistic with continuous data, the chi-square test statistic is calculated by comparing the model of interest to the unrestricted model. An unrestricted model (see Figure 4 for an example with continuous data) is easily estimated with the least squares estimation method. Dependent on the least squares estimation method (i.e., ULS, WLS, DWLS) different test statistics can be calculated: (1) uncorrected, standard chi-square test statistic, (2) mean and variance adjusted test statistic (Satterthwaite type; Satorra, 2000; Satorra & Bentler, 1994), and (3) scaled and shifted test statistic (Asparouhov & Muthén, 2010).

The PML Estimation Method

Model Estimation. Renard (2002) already applied the PML estimation method in the WF approach in a random intercept regression model with binary responses. Similar to the least squares approaches, the PML estimation method can estimate all multilevel models in the WF approach.

Missing Data. Katsikatsou and Moustaki (2017) developed two different approaches to deal with missing data, namely complete-pairs (CP) PML and available-cases (AC) PML. Similarly to the least squares estimation methods for model estimation, the CP procedure only takes the bivariate likelihood into account in case both variables are observed. In the AC procedure, the univariate likelihood of the observed variables where one variable is observed and the other is missing, also contributes to the likelihood. As long as the thresholds are not the parameters of interest, CP and AC provide reliable results and can be used interchangeably (see Katsikatsou & Moustaki, 2017).

Evaluation of Fit and ICC. The PML estimation method can estimate an unrestricted (H1) model for the calculation of the ICC and to provide fit statistics. The PML estimation method can calculate a robust likelihood ratio test (PL-LRT; Katsikatsou & Moustaki, 2016).

The MML Estimation Method

Model Estimation. As the MML estimation method has to integrate out the latent variables, the MML estimation method cannot estimate all models in the WF approach. In practice, the number of latent variables is therefore restricted to a maximum of about four latent variables. The basic number of latent variables in a one factor model (see Figure 2) is already higher than the maximum of four latent variables that the MML estimation method can handle. To model more dimensions, one can turn to Monte Carlo integration solutions instead of using numerical integration. However, this method is very time consuming. The two-tiers MML approach of Cai (2010) may reduce the number of dimensions, but this method still needs many dimensions as many latent variables are related. From a pragmatic point of view, only multilevel regression or path models (without latent variables) can be estimated in the WF approach.

Missing Data. The advantage of the MML estimation method is that it can handle missing data via full information maximum likelihood (Arbuckle, 1996; Neale, 2000).

Evaluation of Fit and ICC. The MML estimation method can fit an unstructured model if the number of items is restricted to about four variables. In practice, it is usually not possible to estimate unrestricted models to calculate the ICCs and fit statistics.

SIMULATION STUDY

A simulation study is performed to evaluate the WF approach under conditions encountered in practice and to compare the WF approach to the well known LF approach with discrete data. In the WF approach, we use the PML estimation method (Jöreskog & Moustaki, 2001) and the diagonally weighted least squares estimation method (DWLS; Muthén et al., 1997), as the latter estimation method has been shown to give accurate results in CFA simulation studies (Beauducel & Herzberg, 2006; Flora & Curran, 2004; Yang-Wallentin, Jöreskog, & Luo, 2010). The estimation methods in the WF approach will be compared to the DWLS and the MML in the LF approach. Using simulated data, we will determine the accuracy and efficiency of the parameter estimates under different conditions. We vary response scales (two-point, four-point), inclusion of a covariate (none, within, between), the number of clusters (200, 1000), and the balancedness of the data (balanced: three units per cluster; unbalanced: 3/6/9 units per cluster). As a consequence of the latter condition, the condition with 200 clusters and unbalanced data includes 90 clusters with 9 units, 70 clusters with 6 units, and 40 clusters with 3 units, while the condition with 1000 clusters and unbalanced data includes 500 clusters with 9 units, 300 clusters with 6 units, and 200 clusters with 3 units.

In a fully crossed design, these factors yield $2 \times 3 \times 2 \times 2 = 24$ different conditions. In addition, 8 conditions were added to include a small misspecification in the first item at the between-level, as this condition is not crossed with conditions including a covariate. In each condition, 500 datasets are generated. The performance of each method is evaluated by calculating the relative bias, namely the accuracy of the estimated parameters (% bias $= (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})/\boldsymbol{\theta}) * 100$) and standard errors (% bias = ((SD-SE)/SD)*100), where SD refers to the standard deviation of the parameter estimates across replications and the SE is the mean of the estimated standard errors across replications.

Data Generation

The data are generated according to the population model as shown in Figure 6. Factor loadings were equal across levels, and more variance was generated at the within-level (i.e., η_{fw}) than at the between-level (i.e., η_{fb}) to mimic real data examples (see Snijders & Bosker, 1999). The residual variances at the within-level imply an identity matrix, to resemble the theta parameterization. The solid lines show the basic multilevel factor model without additional effects of the covariates (i.e., $\mathbf{z_2}$ and $\mathbf{w}$) or a misspecification (i.e., z_1). As we consider a model with zero error variances at the between-level, we assume that factors have the same interpretation across all clusters and ensure that no other variables than the specified latent variables are affecting the between-level responses. This is referred to as measurement invariance restrictions that are necessary to interpret a factor model meaningfully, as described by Muthén (1990), Rabe-Hesketh et al. (2004a), and Mehta and Neale (2005).

Using the parameter values of the basic factor model from Figure 6, the intraclass correlation equals 0.111[1] for all items, which is considered 'common' in, for example, educational data (see Snijders & Bosker, 1999). The

[1] ICC (see Hox et al., 2017) = Var between/(Var between + Var within); = $(.5^2 * 1)/((.5^2 * 1) + (.5^2 \times 4) + 1$ (residual variance in theta parameterization)) = 0.111.

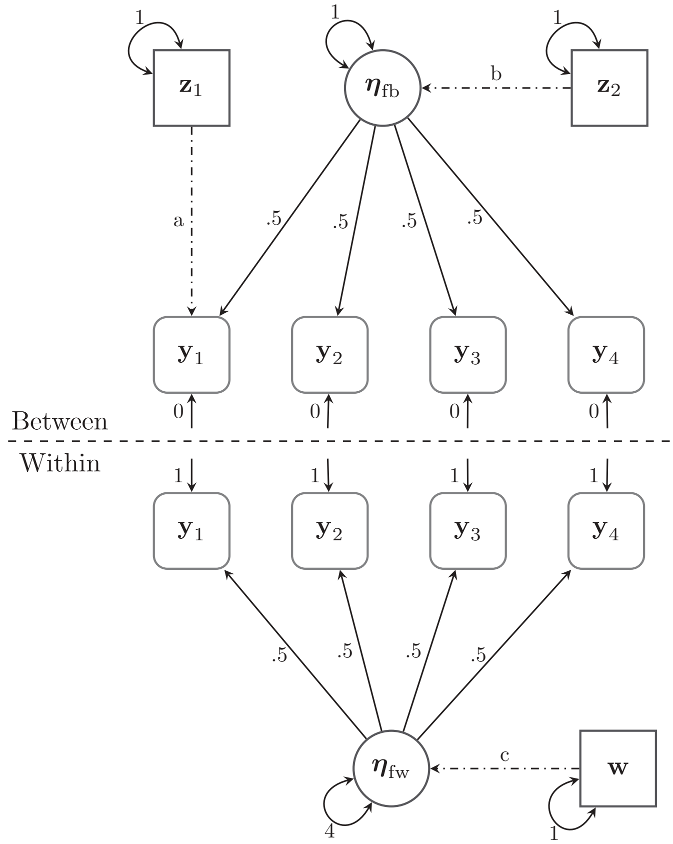

FIGURE 6 A two level data generation model in the LF approach. Note: The solid lines represent the general model and the dashed lines are only added in certain conditions. In a misspecified model a non-zero *a* indicates an additional effect on the first item. Non-zero values for *b* and *c* indicate an additional effect of the covariate on the between- or within-level.

variance on the within-level is chosen to be larger than on the between-level. The factor loadings on both the within- and between-level are chosen to be .5, as these values are commonly used in simulation studies (see Schilling and Bock (2005); Kim, Yoon, Wen, Luo, and Kwok (2015)) or found in empirical data (see Merz and Roesch (2011); Dyer, Hanges, and Hall (2005)).

Continuous data were drawn from a multivariate normal distribution in the WF approach with zero means, and $\boldsymbol{V}_j$ in Equation 22 as the population covariance matrix. Discrete data are obtained by discretizing the continuous data such that the population proportions for the two categories equal .50, .50 and the population proportions for four categories equal approximately .16, .34, .34, and .16.

In some conditions, the basic factor model is extended with additional effects (see the dotted lines in Figure 6). In a misspecified model a non-zero *a* value of .45 has an additional effect on the first item. Non-zero values for *b* and *c* indicate an additional effect of the covariate at the between- or within-level. Both $\mathbf{z}_2$ and $\boldsymbol{w}$ have unity variance and *b* and *c* were set to .3. Appendix E shows a summary of the design factors and the model parameters used in the simulation study.

Estimation

The computer program lavaan (Rosseel, 2012) is used for most calculations. As lavaan does not deal with discrete data in the LF approach yet, we use the computer program Mplus (Muthén and Muthén, 2010). We use robust standard errors for all estimation methods. To obtain comparable results to the LF estimation methods with the probit link, we use the theta parameterization when the WF estimation methods are applied. In agreement with the data generation and according to measurement invariance restrictions, the residuals at the between-level are not estimated. This reduces the number of dimensions drastically and enables us to include the MML estimation method in our simulation study. The MML estimation method is used with adaptive quadrature with 25 integration points per latent variable. The factor variances were fixed to set the metric of the factors. All multilevel models are estimated without restrictions on the factor loadings across levels, as this will give us an indication on how well the parameters are estimated at different levels. Standard missing data procedures are used to deal with unbalanced data and for the PML estimation method we chose the complete pairs (CP) procedure.

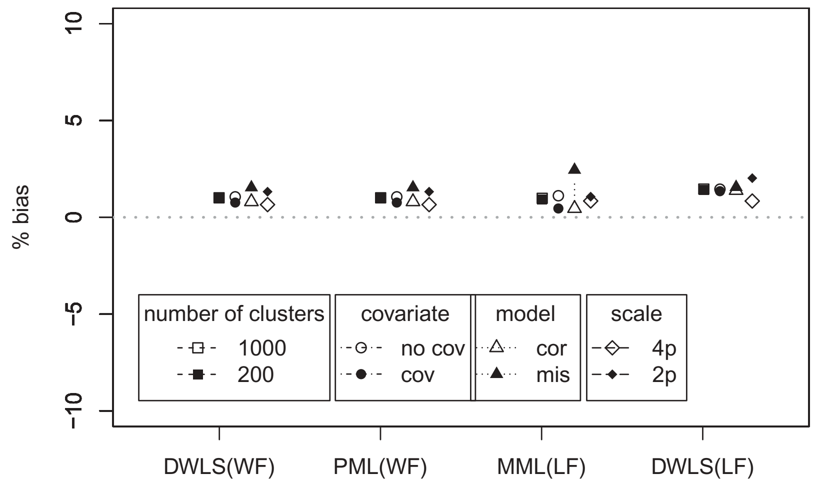

FIGURE 7 Bias of within factor loadings in balanced data.

Results

After applying each of the four estimation methods to each of the 16,000 datasets, we find that the LF DWLS estimation method does not always converge, albeit in a tiny fraction of the number of datasets (0.018 percent of the datasets). To validly compare all estimation methods, the problematic datasets were replaced by new datasets. Below we will describe the relative bias (expressed in percentages) of the estimated factor loadings and the regression coefficients of the covariate at the latent within- or between-factors[2].

Balanced Data

Bias of Factor Loadings. Figure 7 shows the relative bias of the factor loadings at the within-level for different estimation methods. In general, the bias is very low across all conditions and only slightly higher in conditions with a misspecification and the MML estimation method. Figure 8 shows the relative bias of the between-level factor loadings for different estimation methods. All methods show increased bias at the between-level, especially in conditions with a misspecification when the MML estimation method was used. For this specific condition, the relative bias equals 7.12% for two point scales and 11.86% for four point scales.

Bias Standard Errors of Factor Loadings. Figures 9 and 10 show the relative bias of the standard errors associated with the factor loadings at the within- and the between-level, respectively. Figure 9 shows small bias across all conditions with the MML estimation method and the estimation methods in the WF approach. The bias of the standard errors at the between-level (see Figure 10) is small across all estimation methods, except for the MML estimation method where the relative bias is unexpectedly high.

Unbalanced Data

Bias of Factor Loadings. The relative bias of the factor loadings with unbalanced data across all conditions at both the within-level and the between-level generally show the same picture as those with balanced data. Once again the bias in the MML estimation method is somewhat higher in conditions with a misspecification at the within-level and especially high in conditions with a misspecification at the between-level. The corresponding figures are shown in the supplementary materials.

Bias Standard Errors of Factor Loadings. The relative bias of the standard errors associated with the factor loadings at the within-level is shown in Figure 11. The bias is reasonable in the MML estimation method and the WF approaches. Only the DWLS estimation method in the LF approach shows a slightly deviant pattern in many conditions. Figure 12 shows the relative bias of the standard errors associated with the factor loadings at the between-level. The WF estimation methods show higher relative bias across all conditions (i.e., about 10%). The DWLS estimation method in the LF approach shows lower bias than the other estimation methods and the MML estimation method shows more variation in the relative bias across all conditions.

Bias Associated with Regression Coefficients of the Covariate

Overall, the relative bias in both the parameter estimates and the associated standard errors related to the regression coefficients of the covariate is very low across all

[2] All R scripts (e.g., data generation scripts and scripts to analyze the data) are available at the Open Science Foundation (OSF), following the link https://osf.io/fa2gp/.

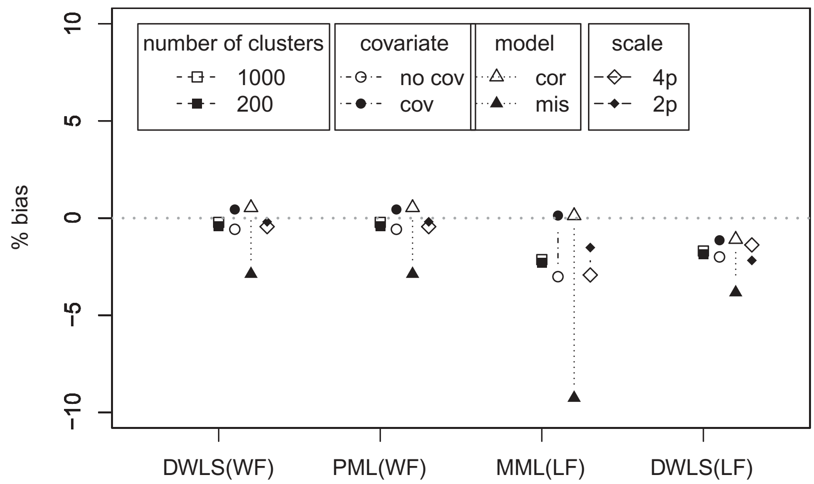

FIGURE 8 Bias of between factor loadings in balanced data.

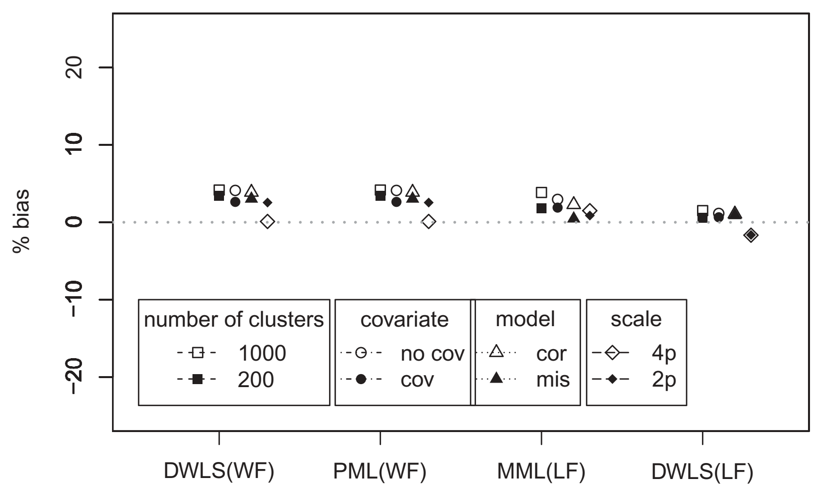

FIGURE 9 SE bias of within factor loadings in balanced data.

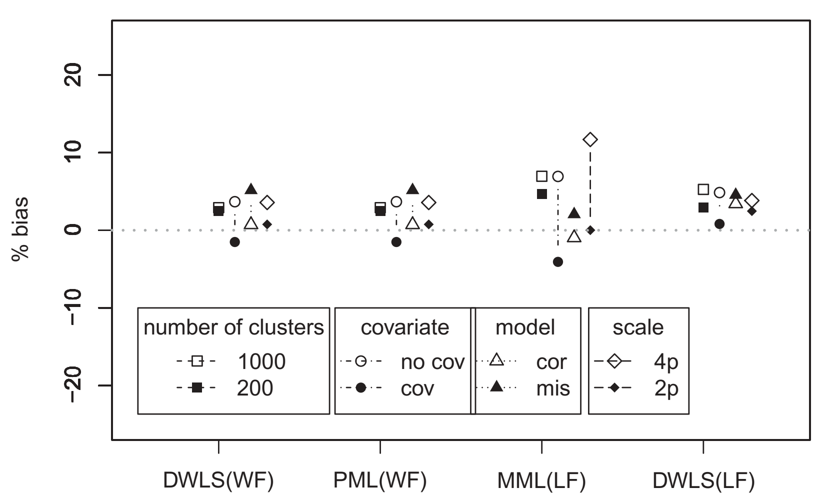

FIGURE 10 SE bias of between factor loadings in balanced data.

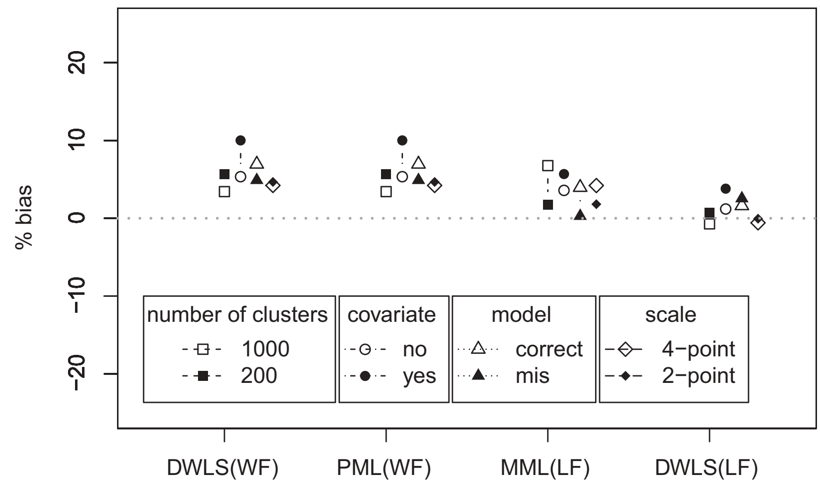

FIGURE 11 SE bias of within factor loadings in unbalanced data.

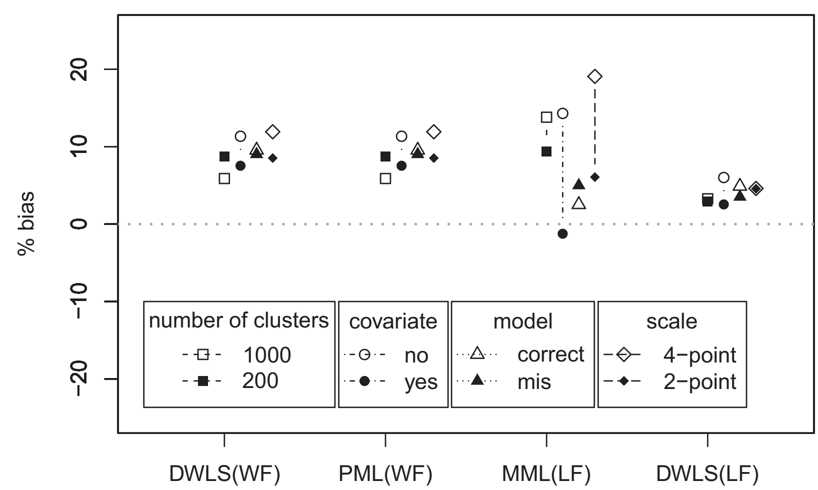

FIGURE 12 SE bias of between factor loadings in unbalanced data.

conditions and all estimation methods. The corresponding figures are shown in the supplementary materials.

Summary of Results

With all estimation methods, the bias of the regression coefficients associated with the covariate and the corresponding standard error at the within- and the between-level is largely unbiased across all conditions. The bias of the factor loadings and the associated standard errors show more variation across estimation methods and/or conditions. Overall, the bias of both the factor loadings and the standard errors across all conditions is higher at the between-level than at the within-level. The WF estimation methods are largely unbiased across conditions and more relative bias of the standard errors at the between-level are noticeable with unbalanced data.The DWLS estimation method in the LF approach shows a deviant pattern compared to all other estimation methods in many conditions. In general, the MML estimation method in the LF approach exhibits more bias than all other estimation methods. Perhaps most prominently the high bias in both the factor loadings and their associated standard errors in conditions with a misspecification.

In describing the parameter estimates, we did not mention the thresholds as these parameters are often not of theoretical interest. For all conditions, the bias in the thresholds was very low for all estimation methods. The figures are shown in the supplementary materials.

ILLUSTRATION

Data

The WF approach to estimate multilevel data is illustrated with data from the dependency scale of a Dutch translation of the Student-Teacher Relationship Scale (STRS; Koomen et al., 2007). As the dataset contains small clusters sizes, the WF approach is very suitable. Through this example, we will show how the WF can be used to estimate multilevel models with different estimation methods. In the dataset of Koomen et al. (2007) primary school teachers filled in six questions regarding overly dependent and clingy child behavior for only a few children (mean = 2.3) from their class, aged 3 to 12:

1. This child fixes his/her attention on me the whole day long.
2. This child reacts strongly to separation from me.
3. This child is overly dependent on me.
4. This child asks for my help when he/she really does not need help.
5. This child expresses hurt or jealousy when I spend time with other children.
6. This child needs to be continually confirmed by me.

A multilevel one factor is specified using the six dependency items (see Figure 6 for a one factor model with 4 items). Responses were given on a 5-point scale ranging from 1 ('definitely does not apply') to 5 ('definitely does apply'). For simplicity, we only selected classes where teachers rated at most three children. This resulted in data from 1047 students (497 boys and 550 girls) that were gathered from 559 primary school teachers (144 men and 415 woman) with an average cluster size of 1.873. In a preliminary analysis a one factor model is fitted, treating the data as continuous, using ML with robust standard errors and fit statistics. Note that with continuous data, the results as well as the fit statistics of the WF approach are identical to the results of the LF approach. This multilevel model fits the data well (χ^2 (18) = 29.990, p = .038, RMSEA = 0.025, CFI = 0.992, SRMR (within-level) = 0.030, SRMR (between-level) = 0.073).

Statistical Analysis

Data were first reordered in the WF approach such that all the children in a class are row-wise displayed. As not all clusters have an equal number of children, clusters with less than three children have missing values (i.e., design missingness see Paragraph 1.3). We will use the same estimation methods as the ones used in the simulation study (i.e., PML and DWLS in the WF approach and MML and DWLS in the LF approach) and mainly focus on the parameter estimates and the associated standard errors. Similar to the simulation study, all WF approaches were estimated with lavaan (Rosseel, 2012) and all LF approaches were estimated with Mplus (Muthén & Muthén, 2010). The gender of the student (i.e., sk) and the gender of the primary school teacher (i.e., sl) are the binary covariates in the analysis.

We will fit the following multilevel models with the STRS data in the WF approach:

- Model 0: Unstructured model without covariates
- Model 1: CFA without covariates
- Model 2: CFA without covariates with measurement invariance restrictions
- Model 3: Unstructured model with binary covariates
- Model 4: CFA with binary covariates
- Model 5: CFA with binary covariates with measurement invariance restrictions
- Model 6: Unstructured model with continuous covariates
- Model 7: CFA with continuous covariates
- Model 8: CFA with continuous covariates with measurement invariance restrictions

All models are fitted according to the proposed steps in section "WF Approach with Discrete Data[3]." We subsequently fit three models without covariates (Model 0 - Model 2), three models with the covariates treated as categorical (Model 3 - Model 5), and three models with covariates treated as continuous (Model 6 - Model 8). Note that estimation methods in the LF approach do not need unstructured models (Model 0 and Model 3).

In models without restrictions (Model 1, Model 4, and Model 7), the factor loadings at the within-level can be different from the factor loadings at the between-level and the residuals at the between-level (i.e., random intercept variances leftover) are freely estimated. These kind of multilevel models have a complicated interpretation as the within-level factor loadings should then be interpreted as a summary of all possible cluster factor loadings. Nevertheless, the model without restrictions can serve as a baseline model for models with restrictions across clusters. Models with measurement invariance restrictions (Model 2, Model 5, and Model 8) ensure that factors have the same interpretation across all clusters, and restrictions on the residuals at the between-level ensure that the clusters have the same intercepts and that no other variables than the specified latent variables are affecting the between-level responses (see Mehta & Neale, 2005; Muthén, 1990; Rabe-Hesketh et al., 2004a).

It is important to note that the MML estimation method cannot estimate Model 0 or CFA models without measurement invariance restrictions, as the number of latent variables in all other models exceeds the number of dimensions the MML estimation method can handle. This has to do

[3] The syntax can be found at the OSF. The syntax used to fit the STRS data is quite similar to the syntax printed in Appendix D (see Appendix B for adding covariates).

with the estimation of the errors on the between-level, that are modeled as additional latent variables.

Results

To briefly illustrate the use of fit statistics in the WF approach, Table 1 contains the intra-class coefficients (ICC) according to the DWLS and PML estimation methods in the WF approach and the DWLS estimation method in the LF approach. The ICC values are almost equivalent across estimation methods. Forcing the MML to calculate the ICC will lead to too many latent variables which is computationally too intensive. Table 2 shows the χ^2 values and the degrees of freedom of the models of interest and the unstructured models for the STRS data using the DWLS estimation method with the scaled and shifted test statistic (Asparouhov & Muthén, 2010). In the WF approach, the model of interest and the unrestricted model are necessary to obtain the multilevel test statistic. The last two columns indicate the chi-square test statistics and the associated degrees of freedom for the various multilevel models obtained via a scaled chi-square difference test (Satorra, 2000). The results indicate that the general model with or without covariates fits the data well, while models with invariance restrictions do not fit the data. To give a general idea of the estimation time of the used estimation methods, we saved the estimation time of a model without covariates and errors on the between level; DWLS in the WF approach: 0.11 minutes, PML in the WF approach: 3.24 minutes, MML in the long format approach: 4:25 minutes, and DWLS in the LF approach: 1:53 minutes.

Column 2 to 5 of Table 3 contains the parameter estimates and the standard errors of a model without measurement invariance restrictions across levels for different estimation methods with a discrete covariate. Only the theta parameterization ensures that the WF approach is comparable to the LF approaches. As the DWLS estimation method in the LF approach cannot include the covariate as a discrete covariate, Column 6 to 11 contain the parameter estimates and the standard errors of a model without measurement invariance restrictions across levels for different estimation methods with a continuous covariate. During estimation we noticed that the DWLS in the WF approach had problems constructing a weight matrix (i.e., **W**), which is most likely due to treating the binary covariates as continuous. The model shown in Table 3 contains too many dimensions (i.e., six errors variances at the between-level and two latent variables) to estimate this model with the MML estimation method in the LF approach.

TABLE 1
ICC for Different Estimation Methods

	DWLS(LF)	*DWLS(WF)*	*PML(WF)*
Item 1	0.463	0.460	0.453
Item 2	0.362	0.409	0.379
Item 3	0.145	0.151	0.137
Item 4	0.274	0.276	0.283
Item 5	0.421	0.417	0.412
Item 6	0.203	0.207	0.204

TABLE 2
χ^2 Test Statistics for the DWLS Estimation Method in the WF

	H1		*H0*		*model fit*	
	df	χ^2	*df*	χ^2	*df*	χ^2
Models without covariates						
Model 0 & Model 1:	164	152.111	182	166.752	18	8.010
Model 0 & Model 2:	164	152.111	193	310.396	29	112.336*
Models with binary covariates						
Model 3 & Model 4:	226	212.385	254	238.925	28	20.783
Model 3 & Model 5:	226	212.385	265	389.558	39	135.324*
Models with covariates						
Model 6 & Model 7:	234	168.540	262	192.628	28	19.000
Model 6 & Model 8:	234	168.540	273	342.596	39	133.402*

Note: * = $p < 0.05$

The results of Table 3 indicate that the parameter estimates of the PML estimation method and the DWLS estimation method in the WF approach are quite similar and only the parameter estimates for the DWLS estimation method in the LF approach are somewhat deviating. For all estimation methods, the error variances are not equal to zero, which indicate that the intercepts of all classes are not equal to each other and that other variables than the specified latent variables may affect the between-level responses.

Column 2 to 7 of Table 4 show the parameter estimates and the standard errors of the STSR models with measurement invariance restrictions across levels. The model fits poor in terms of the χ^2 fit statistic, however, the calculation of the Correlation Root Mean Squared Residual (CRMR; Bentler, 1985; Maydeu-Olivares, 2017; Ogasawara, 2001), that is less dependent on the sample size, showed reasonable fit values (e.g., 0.0148, $z = 1.425$, $p > .05$ for Model 2). The measurement invariance restrictions facilitate the interpretation of the effect of the covariates. As boys were scored 0 and girls 1, the unstandardized effect of dependency on the gender of the child equals 0.106 in, for example, the WF DWLS estimation method with a discrete covariate. This means that teachers experience more dependency with girls than with boys. The unstandardized direct effect in the WF DWLS of dependency on the gender of the teacher is −0.098 indicating that male teachers experience more dependency than female teachers. Overall, the parameter estimates of the multilevel MML are quite similar to the ones obtained by the WF approach. The effect of the covariate is also about equal in terms of ratios. The parameter estimates of the LF DWLS are again somewhat deviating. The last six columns of Table 4 show the parameter estimates and the standard errors of the models with and without measurement invariance restrictions with continuous covariates in the WF approach.

TABLE 3
Parameter Estimates of Model 4 and Model 7

	binary covariate				*continuous covariate*					
	DWLS(WF)		*PML(WF)*		*DWLS(WF)*		*PML(WF)*		*DWLS(LF)*	
	$\boldsymbol{\theta}$	*SE*	Var	*SE*	Var	*SE*	Var	*SE*	$.5^2 * 1$	*SE*
$\lambda^w_{1,1}$	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
$\lambda^w_{1,2}$	2.475	0.446	2.352	0.426	2.477	0.446	2.361	0.434	2.029	0.319
$\lambda^w_{1,3}$	2.058	0.354	1.947	0.364	2.008	0.340	1.951	0.367	1.589	0.242
$\lambda^w_{1,4}$	1.428	0.247	1.445	0.257	1.392	0.238	1.445	0.258	1.166	0.177
$\lambda^w_{1,5}$	1.889	0.307	1.864	0.322	1.874	0.302	1.862	0.325	1.574	0.222
$\lambda^w_{1,6}$	3.097	0.592	3.069	0.679	3.161	0.618	3.053	0.681	2.186	0.333
Var(η^w)	0.217	0.064	0.221	0.069	0.222	0.065	0.223	0.070	0.325	0.077
Var(sk)	-	-	-	-	0.249	0.003	0.250	0.001	-	-
Int(sk)	-	-	-	-	0.535	0.021	0.521	0.012	-	-
η^w on sk	0.073	0.026	0.070	0.026	0.108	0.054	0.112	0.042	0.141	0.050
$\lambda^b_{1,1}$	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
$\lambda^b_{1,2}$	1.259	0.207	1.227	0.222	1.264	0.209	1.229	0.227	1.281	0.230
$\lambda^b_{1,3}$	0.517	0.115	0.503	0.117	0.519	0.114	0.503	0.117	0.544	0.119
$\lambda^b_{1,4}$	0.835	0.164	0.833	0.164	0.840	0.164	0.833	0.164	0.892	0.183
$\lambda^b_{1,5}$	0.934	0.126	0.965	0.137	0.936	0.127	0.965	0.142	1.034	0.169
$\lambda^b_{1,6}$	0.608	0.129	0.610	0.153	0.613	0.133	0.609	0.156	0.652	0.139
Var(η^b)	0.749	0.172	0.689	0.156	0.761	0.173	0.701	0.160	0.658	0.170
η^b on sl	−0.174	0.078	−0.163	0.077	−0.287	0.131	−0.271	0.128	−0.248	0.124
Var(sl^b)	-	-	-	-	0.191	0.009	0.195	0.007	-	-
Int(sl)	-	-	-	-	0.742	0.007	0.734	0.015	-	-
$\Theta^b_{1,1}$	0.261	0.143	0.304	0.132	0.267	0.143	0.304	0.133	0.483	0.138
$\Theta^b_{2,2}$	0.395	0.206	0.299	0.186	0.404	0.210	0.306	0.188	0.244	0.157
$\Theta^b_{3,3}$	0.138	0.114	0.115	0.097	0.131	0.111	0.116	0.098	0.113	0.095
$\Theta^b_{4,4}$	0.011	0.111	0.083	0.116	0.001	0.109	0.083	0.116	0.021	0.107
$\Theta^b_{5,5}$	0.604	0.188	0.587	0.182	0.603	0.187	0.585	0.180	0.611	0.138
$\Theta^b_{6,6}$	0.530	0.233	0.534	0.239	0.557	0.248	0.527	0.237	0.370	0.129

Note: (1) Thresholds are not shown here; (2) For identification purposes the first factor loading on the between-level and the within-level are fixed to one.

Conclusion

The STRS data shows that the WF approach can be used in datasets with a relatively small number of units in each cluster. Using the stepwise approach from section "Multilevel SEM in the WF Approach" and section "WF Approach with Discrete Data", various CFA models, including models with covariates and measurement bias restrictions, can be fitted to the STRS data. The DWLS in both the LF- and WF approch and the PML estimation method can estimate all suggested models. Due to the increasing dimensionality, the MML estimation method cannot fit models with error variances on the between-level (see Table 3). We hesitate to interpret the results of the the DWLS estimation method in the LF approach, as it shows a different pattern of parameter estimates compared to all other estimation methods.

Discussion

In this article, we united the fragmented literature on the WF approach into a general framework for random intercept SEM multilevel models for a small number of observations in a cluster that can handle unbalanced data, missing data, and covariates at both the within- and the between-level. In addition, we also extent the WF approach for discrete data and provide an illustration. With small cluster sizes, the least squares estimation methods and the PML estimation method are very useful to fit all multilevel models in the WF approach. Only the MML estimation method is harder to use in the WF approach as it is computationally very intensive and only of practical use in the estimation of multilevel regression or path models. To make the WF approach applicable, we used the Open Science Framework to provide the readers with the R script of the empirical example, and the R scripts and artificial data that corresponds to the models presented in the appendix.

In general, modeling multilevel data in the WF approach has several advantages over modeling multilevel data in the LF approach. First, the WF approach is computationally very efficient compared to the MML estimation method in

TABLE 4
Parameter Estimates of Model 5 and Model 8

	binary covariate						continuous covariate					
	DWLS(WF)		PML(WF)		MML(LF)		DWLS(WF)		PML(WF)		DSMV(LF)	
	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE
$\lambda_{1,1}$	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
$\lambda_{1,2}$	1.611	0.167	1.697	0.204	1.845	0.186	1.613	0.168	1.697	0.209	1.881	0.256
$\lambda_{1,3}$	0.994	0.103	0.993	0.126	1.174	0.117	0.989	0.102	0.993	0.127	1.206	0.151
$\lambda_{1,4}$	1.115	0.126	1.161	0.142	1.246	0.129	1.109	0.125	1.161	0.140	1.109	0.151
$\lambda_{1,5}$	1.138	0.101	1.224	0.121	1.300	0.120	1.136	0.100	1.224	0.124	1.469	0.173
$\lambda_{1,6}$	1.081	0.105	1.100	0.139	1.318	0.130	1.080	0.105	1.101	0.141	1.554	0.175
Var(η^w)	0.409	0.070	0.374	0.062	0.324	0.052	0.415	0.070	0.377	0.064	0.440	0.086
Var(sk)	-	-	-	-	-	-	0.249	0.003	0.250	0.001	-	-
Int(sk)	-	-	-	-	-	-	0.535	0.021	0.521	0.012	-	-
η^w on sk	0.106	0.037	0.095	0.035	0.143	0.050	0.157	0.082	0.152	0.056	0.164	0.057
Var(η^b)	0.346	0.072	0.289	0.068	0.266	0.053	0.351	0.073	0.293	0.070	0.267	0.065
η^b on sl	−0.098	0.052	−0.089	0.049	−0.119	0.072	−0.162	0.088	−0.148	0.081	−0.127	0.078
Var(sl^b)	-	-	-	-	-	-	0.191	0.009	0.195	0.007	-	-
Int(sl)	-	-	-	-	-	-	0.742	0.007	0.734	0.015	-	-

Note:(1) Thresholds are not shown here; (2) For identification purposes the first factor loading on the between-level and the within-level are fixed to one.

the LF approach opening up the possibility to analyze models with many latent variables. The illustration showed that a simple one factor model with errors on the between-level contains already too many latent variables to estimate with the MML estimation method in the LF approach. Estimating this model in the WF approach with PML and DWLS causes no problems. This offers new possibilities to estimate factor models with errors on the between-level. Second, the WF approach does not need as many restrictions as the LF approach. Where the LF approach implicitly forces equality constraints on, for example, the factor loadings and the thresholds across clusters, the WF approach can test all these restrictions. Third, there is no special SEM software needed to perform multilevel analysis with discrete data, as single level software is capable of performing the multilevel analysis. Despite the potential advantages, there are of course a number of limitations. First, larger cluster sizes and/or more variables require a larger number of clusters. Second, highly unbalanced data are problematic in the WF approach. If, for example, all clusters contain no more than five units, except for one cluster with twenty units, we would advice to randomly sample five units from this cluster to deal with this highly unbalanced data in the WF approach. Third, it is tedious to specify a multilevel model in the WF approach using current SEM software. As the maximum number of units increase, the size of the model syntax increases too. In the future, we plan to develop scripts that will automatically generate the model syntax.

The simulation study showed that the estimation methods in the WF approach are capable to estimate multilevel models under the simulated conditions. The parameter estimates of the estimation methods in this simulation study with the WF approach are largely unbiased and not so much influenced by the balancedness of the data, the number of clusters, and model misspecification. It is noticeable that the bias of the parameter estimates and the standard errors at the between-level are larger than the bias at the within-level. Of course, the between-level contains less units than the within-level, which influences the bias of the parameter estimates and the standard errors. The DWLS and the PML estimation method in the WF approach show very similar results, which can be explained by the fact that they both rely on bivariate information. The MML estimation method in the LF approach shows comparable results to the estimation methods in the WF approach, except in conditions with a misspecification, where the bias in the parameter estimates is alarmingly high. The high bias of the standard errors at the between-level with four response scales in the MML estimation method, was also surprising. The DWLS estimation method in the LF approach is hindered by some convergence problems and shows a different pattern of results compared to all other estimation methods. As it is impossible to consider all plausible scenarios in a single simulation study, generalization beyond the range of the conditions studied should be undertaken with caution.

The illustrative dataset on the student-teacher relationship shows that the WF approach is capable of estimating different multilevel models in real data. If the MML estimation is feasible (as in Table 4), the results of the WF approach are very similar to the results of the MML estimation method in the LF approach. The results of the DWLS in the LF approach are typically somewhat deviant. Although we did not encounter difficulties with empty cells in our example, empty cells caused by within-cluster missingness or design missingness can be problematic in the WF approach. Potential solutions for datasets with empty cells are discussed by Agresti and Yang (1987).

In the illustration, we briefly showed how to calculate the χ^2 goodness-of-fit statistic in the WF approach. As the fit in a multilevel model expresses the combined (mis)fit at both the within- and the between-level, more attention should be given to this topic. The within-level has generally more influence on the overall fit than the between-level, as the number of units at the within-level is higher than the number of units at the between-level. Ryu and West (2009) and Boulton (2011) proposed level specific fit measures for multilevel SEM.

In this study, we only considered models with a random intercept. Multilevel models can be extended by adding random slopes, where the impact of covariates is allowed to vary across clusters. The estimation of random slopes requires case-wise estimation. With discrete data in the LF, only the MML estimation method can estimate models with a random slope. In the WF approach, the PML estimation method seems best suited to estimate models including a random slope in for example generalized linear mixed models (see Bellio & Varin, 2005; Cho & Rabe-Hesketh, 2011; Tibaldi et al., 2007). Compared to the MML in the LF approach, the PML estimation method in the WF approach can in theory estimate many random slopes and other latent variables. The WLS estimation method in the WF approach uses a two-step estimation procedure and can therefore not estimate models with random slopes.

The PML estimation method seems a promising, perhaps underused, estimation method that was initially developed for discrete data (see Jöreskog & Moustaki, 2001). So far, exogenous covariates in the PML estimation method are regressed out first and further calculations are done on the residual correlations. However, in the WF approach the covariates are considered as stochastic variables, where the PML estimation method has to deal with pairs of variables that are potentially a mixture of discrete and continuous variables and combinations thereof. This is the first study that deals with a mixture of discrete and continuous variables in estimating SEM with the PML estimation method. As mentioned above, the PML estimation method is also capable of estimating random slopes. In addition, with very unequal cluster sizes, the PML estimation method could be extended by using weights (e.g., Joe & Lee, 2009; Renard, 2002).

In future research, the WF approach can easily be extended to more than two levels or used to estimate multi-group multilevel models or multilevel longitudinal models, where the mean structure is also of importance. Additionally, the possibilities of the PML estimation method with case-wise estimation can be explored to estimate models with random slopes and/or models in the WF approach where the number of columns (variables $\times$ observations) is larger than the number of rows (number of clusters).

ACKNOWLEDGEMENTS

The authors thank Helma Koomen of the University of Amsterdam for making her data available for secondary analysis. We are also thankful to Karin Schermelleh-Engel and Wen Wei Loh for helpful comments on an earlier draft of this article.

FUNDING

This publication is supported by Special Research Fund (BOF) open competition grant from Ghent University.

ORCID

M.T. Barendse http://orcid.org/0000-0002-1632-0303
Y. Rosseel http://orcid.org/0000-0002-4129-4477

REFERENCES

Agresti, A., & Yang, M.-C. (1987). An empirical investigation of some effects of sparseness in contingency tables. *Computational Statistics & Data Analysis*, *5*, 9–21. doi:10.1016/0167-9473(87)90003-X

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. In Marcoulides & R. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Erlbaum.

Asparouhov, T., & Muthén, B. (2007, August). Computationally efficient estimation of multilevel high-dimensional latent variable models. In *proceedings of the 2007 JSM meeting in Salt Lake City, Utah, Section on Statistics in Epidemiology* (pp. 2531–2535).

Asparouhov, T., & Muthén, B. (2010). Simple second order chi-square correction. *Mplus technical appendix*.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, *28*, 135–167. doi:10.3102/10769986028002135

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in cfa. *Structural Equation Modeling*, *13*, 186–203. doi:10.1207/s15328007sem1302_2

Bellio, R., & Varin, C. (2005). A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling*, *5*, 217–227. doi:10.1191/1471082X05st095oa

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In PR Krishnaiah (Ed.), *Multivariate Analysis* (Vol. 6, pp.9–42)..

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an em algorithm. *Psychometrika*, *46*, 443–459. doi:10.1007/BF02293801

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons.

Boulton, A. J. (2011). *Fit index sensitivity in multilevel structural equation modeling* (PhD thesis). University of Kansas.

Browne, M. W. (1984). Asymptotically distribution–Free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83. doi:10.1111/bmsp.1984.37.issue-1

Browne, M. W. (1995). Covariance structures. In *Topics in applied multivariate analysis* (pages 72–141). ID M. Hawkins. Cambridge, UK: Cambridge University Press.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147–158. doi:10.1037/0033-2909.101.1.147

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, *75*, 581–612. doi:10.1007/s11336-010-9178-0

Cho, S.-J., & Rabe-Hesketh, S. (2011). Alternating imputation posterior estimation of models with crossed random effects. *Computational Statistics & Data Analysis*, *55*, 12–25. doi:10.1016/j.csda.2010.04.015

Chou, C.-P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. *Structural Equation Modeling*, *5*, 247–266. doi:10.1080/10705519809540104

Croon, M. A., & van Veldhoven, M. J. (2007). Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods*, *12*, 45. doi:10.1037/1082-989X.12.1.45

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*, 529–569. doi:10.1207/s15327906mbr3804_5

Depaoli, S., & Clifton, J. P. (2015). A bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling*, *22*, 327–351. doi:10.1080/10705511.2014.937849

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, *16*, 149–167. doi:10.1016/j.leaqua.2004.09.009

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491. doi:10.1037/1082-989X.9.4.466

Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*. New York, NY: Springer Science & Business Media.

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, *53*, 455–467. doi:10.1007/BF02294400

Guo, B., Perron, B. E., & Gillespie, D. F. (2008). A systematic review of structural equation modelling in social work research. *British Journal of Social Work*, *39*, 1556–1574. doi:10.1093/bjsw/bcn101

Hedeker, D., & Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 933–944. doi:10.2307/2533433

Holtmann, J., Koch, T., Lochner, K., & Eid, M. (2016). A comparison of ml, wlsmv, and bayesian methods for multilevel structural equation models in small samples: A simulation study. *Multivariate Behavioral Research*, *51*, 661–680. doi:10.1080/00273171.2016.1208074

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Joe, H., & Lee, Y. (2009). On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, *100*, 670–685. doi:10.1016/j.jmva.2008.07.004

Jöreskog, K. G. (1981). Analysis of covariance structures. *Scandinavian Journal of Statistics*, *8*, 65–92.

Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, *36*, 347–387. doi:10.1207/S15327906347-387

Katsikatsou, M., & Moustaki, I. (2016). Pairwise likelihood ratio tests and model selection criteria for structural equation models with ordinal variables. *Psychometrika*, *81*, 1046–1068. doi:10.1007/s11336-016-9523-z

Katsikatsou, M., & Moustaki, I. (2017). Pairwise likelihood estimation for confirmatory factor analysis models with ordinal variables and data that are missing at random. *Submitted to Journal of Royal Statistical Society Series C*.

Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., & Jöreskog, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics & Data Analysis*, *56*, 4243–4258. doi:10.1016/j.csda.2012.04.010

Kim, E. S., Yoon, M., Wen, Y., Luo, W., & Kwok, O.-M. (2015). Within-level group factorial invariance with multilevel data: Multilevel factor mixture and multilevel mimic models. *Structural Equation Modeling*, *22*, 603–616. doi:10.1080/10705511.2014.938217

Koomen, H., Verschueren, K., & Pianta, R. C. (2007). *Leerling leerkracht relatie vragenlijst- handleiding g. [Student–Teacher Relationship Scale: Manual]*. Houten: Bohn Stafleu van Loghum.

Koomen, H. M., Verschueren, K., van Schooten, E., Jak, S., & Pianta, R. C. (2012). Validating the student-teacher relationship scale: Testing factor structure and measurement invariance across child gender and age in a dutch sample. *Journal of School Psychology*, *50*, 215–234. doi:10.1016/j.jsp.2011.09.001

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 pisa 2003 participating countries. *Learning and Individual Differences*, *19*, 355–365. doi:10.1016/j.lindif.2008.10.009

Li, H., Fortner, C. K., & Lei, X. (2015). Relationships between the use of test results and us students' academic performance. *School Effectiveness and School Improvement*, *26*, 258–278. doi:10.1080/09243453.2014.898662

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, *80*, 221–239.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203. doi:10.1037/a0012869

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–226. doi:10.1146/annurev.psych.51.1.201

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, *32*, 215–253. doi:10.1207/s15327906mbr3203_1

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika*, *82*, 533–558. doi:10.1007/s11336-016-9552-7

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, pages 110–133. doi:10.2307/1130295

McDonald, R. P. (1993). A general model for two-level data with responses missing at random. *Psychometrika*, *58*, 575–585. doi:10.1007/BF02294828

McDonald, R. P., & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology*, *42*, 215–232. doi:10.1111/bmsp.1989.42.issue-2

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*, 259–283. doi:10.1037/1082-989X.10.3.259

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122. doi:10.1007/BF02294746

Merz, E. L., & Roesch, S. C. (2011). Modeling trait and state variation using multilevel factor analysis with panas daily diary data. *Journal of Research in Personality*, *45*, 2–9. doi:10.1016/j.jrp.2010.11.003

Muthén, B., Du Toit, S. H., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika*, *75*, 1–45.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557–585.

Muthén, B. O. (1990). *Mean and covariance structure analysis of hierarchical data* (Vol. 62, pp. 1–64). Los Angeles, LA: University of California: Department of Statistics (UCLA Statistics Series).

Muthén, B. O., & Satorra, A. (1995). Technical aspects of Muthén’s liscomp approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, *60*, 489–503.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus: Statistical analysis with latent variables: User’s guide*. Los Angeles, CA: Muthén &Muthén.

Neale, M. C. (2000). Modeling longitudinal and multilevel data: Practical issues, applied approaches and specific examples. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Individual fit, heterogeneity, and missing data in multigroup sem* (pp. 249–267). Hillsdale, NJ: Erlbaum.

NLSAH. (2005). *National longitudinal study of adolescent health*. Retrieved from http//:www.cpc.unc.edu/addhealth

Ogasawara, H. (2001). Standard errors of fit indices using residuals in structural equation modeling. *Psychometrika*, *66*, 421–436. doi:10.1007/BF02294443

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*, 443–460. doi:10.1007/BF02296207

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika*, *47*, 337–347. doi:10.1007/BF02294164

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004a). Generalized multilevel structural equation modeling. *Psychometrika*, *69*, 167–190. doi:10.1007/BF02295939

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004b). Gllamm manual. uc berkeley division of biostatistics (working paper series 1160). University of California, Berkeley.

Renard, D. (2002). *Topics in modeling multilevel and longitudinal data* (Doctoral dissertation). Belgium.

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

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*, *16*, 583–601. doi:10.1080/10705510903203466

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In *Innovations in multivariate statistical analysis* (pp. pages 233–247). Boston, MA: Springer.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In Clogg and Clogg (Ed.), *Latent variables analysis: Applications for developmental research* (pp.399–419). Thousand Oaks, CA: Sage.

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, *70*, 533–555.

Schmidt, W. H. (1969). *Covariance structure analysis of the multivariate random effects model* (PhD thesis). University of Chicago, Department of Education.

Snijders, T., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Tibaldi, F. S., Verbeke, G., Molenberghs, G., Renard, D., Noortgate, W., & Boeck, P. (2007). Conditional mixed models with crossed random effects. *British Journal of Mathematical and Statistical Psychology*, *60*, 351–365. doi:10.1348/000711006X110562

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, *92*, 1–28. doi:10.1007/s10182-008-0060-7

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116*, 363. doi:10.1037/0033-2909.116.2.363

Wirth, R., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58. doi:10.1037/1082-989X.12.1.58

Yang, Y. (2018). Structural equation modelling. In P. Brough (Ed.), *Advanced research methods for applied psychology* (pp. 258–270). New York, NY: Routledge.

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, *17*, 392–423. doi:10.1080/10705511.2010.489003

APPENDIX

Appendix A: lavaan syntax for a multilevel one factor model with a covariate on both levels (i.e., *w* and *z*) in the WF approach

```
# Syntax corresponding to Figure 2 with three observations in each cluster model <- '
  ### Step 2: Model for the within level
  ### Step 3: Equality constraints on the within level (labels e to h =
  ### within factor loadings (λ), fvw1 to fvw3 = within factors (ηf w),
  ### and rw1 to rw4 = residuals variances (Θ))
  # within structure via factor loadings and within factors variances
  fw1 =~ e*y1.1 + f*y2.1 + g*y3.1 + h*y4.1
  fw2 =~ e*y1.2 + f*y2.2 + g*y3.2 + h*y4.2
  fw3 =~ e*y1.3 + f*y2.3 + g*y3.3 + h*y4.3
  fw1 ~~ fvw*fw1
  fw2 ~~ fvw*fw2
  fw3 ~~ fvw*fw3

  # uncorrelated fw1, fw2, fw3
  fw1 ~~ 0*fw2 + 0*fw3; fw2 ~~ 0*fw3

  # within intercepts (fixed to zero)
  y1.1 + y2.1 + y3.1 + y4.1 ~ 0*1
  y1.2 + y2.2 + y3.2 + y4.2 ~ 0*1
  y1.3 + y2.3 + y3.3 + y4.3 ~ 0*1

  # common residual variances
  y1.1 ~~ rw1*y1.1; y1.2 ~~ rw1*y1.2; y1.3 ~~ rw1*y1.3
  y2.1 ~~ rw2*y2.1; y2.2 ~~ rw2*y2.2; y2.3 ~~ rw2*y2.3
  y3.1 ~~ rw3*y3.1; y3.2 ~~ rw3*y3.2; y3.3 ~~ rw3*y3.3
  y4.1 ~~ rw4*y4.1; y4.2 ~~ rw4*y4.2; y4.3 ~~ rw4*y4.3

  ### Step 4: construct latent variables that represent random intercepts
  ### of the variables y1,y2,y3, and y4 on the between level
  by1 =~ 1*y1.1 + 1*y1.2 + 1*y1.3
  by2 =~ 1*y2.1 + 1*y2.2 + 1*y2.3
  by3 =~ 1*y3.1 + 1*y3.2 + 1*y3.3
  by4 =~ 1*y4.1 + 1*y4.2 + 1*y4.3

  # between intercepts
  by1 + by2 + by3 + by4 ~ 1

  ### Step 5: construct a model at the between level
  # between factor
  fb =~ a*by1 + b*by2 + c*by3 + d*by4

  # residual variances on the between-level
  by1 ~~ by1; by2 ~~ by2; by3 ~~ by3; by4 ~~ by4

  # between factor not correlated with the within factors fb ~~ 0*fw1 + 0*fw2 + 0*fw3
'

# fitting the model with ML
fit <- sem(model, data = wideData)
summary(fit)
```

Appendix B: additional lavaan syntax to add a covariate on both the witin-level (i.e., w) and the between-level (i.e., z) in the model of Appendix A

```
#within covariate (w) part: common variance (wwv)
w.1 ~~ wwv*w.1; w.2 ~~ wwv*w.2; w.3 ~~ wwv*w.3;

# common mean for w
w.1 + w.2 + w.3 ~ iw*1

# common slope (ws)
fw1 ~ ws*w.1
fw2 ~ ws*w.2
fw3 ~ ws*w.3

# insert a covariate on the between-level z (as latent variable)
zb =~ 1*z; z ~ 0*1; z ~~ 0*z
zb ~~ zb
zb ~ 1
zb ~~ 0*fw1 + 0*fw2 + 0*fw3

# between regression
fb ~ breg*zb
```

Appendix C: lavaan syntax for a multilevel regression model with discrete responses in the WF approach

```
# Random intercept regression model for discrete outcomes (4-point scales) model <- '
  ### Step 2: Model for the within level
  ### Step 3: Equality constraints on the within level: not necessary
  # common plain variance for y
  y1 ~~ 1*y1; y2 ~~ 1*y2; y3 ~~ 1*y3; y4 ~~ 1*y4;

  ### Step 4: construct a latent variable that represents the random
  ### intercept of y on the between level and put equality constraints
  ### on the th1 to th3 = thresholds (τ))

  # construct beween version of y
  yb =~ 1*y1 + 1*y2 + 1*y3 + 1*y4
  yb ~ 0*1

  # equal thresholds
  y1 + y2 + y3 + y4 | th1*t1 + th2*t2 + th3*t3

  ### Step 5: construct a model at the between level yb ~~ bvar*yb
'

# fit the model with for example the DWLS estimation method Wfit <- sem(model, data = D.wide,
ordered = paste0("y",1:4),
  estimator = "DWLS", parameterization = "theta") summary(Wfit)
```

Appendix D: lavaan syntax for the multilevel one factor model with discrete responses in the WF approach

```
# Syntax corresponding to three observations in each cluster
# with discrete outcomes (4-point scales) in the theta parameterization model <- '

  ### Step 2: Model for the within level
```

```
  ### Step 3: Equality constraints on the within level (labels e to h =
  ### within factor loadings (λ) and fvw1 to fvw3 = within factors (ηf w))
  # within structure via factor loadings, common variance within factors
  fw1 =~ e*y1.1 + f*y2.1 + g*y3.1 + h*y4.1
  fw2 =~ e*y1.2 + f*y2.2 + g*y3.2 + h*y4.2
  fw3 =~ e*y1.3 + f*y2.3 + g*y3.3 + h*y4.3
  fw1 ~~ fvw*fw1
  fw2 ~~ fvw*fw2
  fw3 ~~ fvw*fw3
  # uncorrelated fw1, fw2, fw3
  fw1 ~~ 0*fw2 + 0*fw3; fw2 ~~ 0*fw3

  # within intercepts (fixed to zero)
  # y1.1 + y2.1 + y3.1 + y4.1 ~ 0*1)
  # y1.2 + y2.2 + y3.2 + y4.2 ~ 0*1)
  # y1.3 + y2.3 + y3.3 + y4.3 ~ 0*1)

  # unit residual variances: theta parameterization
  y1.1 ~~ 1*y1.1; y1.2 ~~ 1*y1.2; y1.3 ~~ 1*y1.3
  y2.1 ~~ 1*y2.1; y2.2 ~~ 1*y2.2; y2.3 ~~ 1*y2.3
  y3.1 ~~ 1*y3.1; y3.2 ~~ 1*y3.2; y3.3 ~~ 1*y3.3
  y4.1 ~~ 1*y4.1; y4.2 ~~ 1*y4.2; y4.3 ~~ 1*y4.3

  ### Step 4: construct latent variables that represent random intercepts
  ### with corresponding restrictions on the th1 to th3 = thresholds (τ))

  # between version of y1,y2,y3,y4
  by1 =~ 1*y1.1 + 1*y1.2 + 1*y1.3
  by2 =~ 1*y2.1 + 1*y2.2 + 1*y2.3
  by3 =~ 1*y3.1 + 1*y3.2 + 1*y3.3
  by4 =~ 1*y4.1 + 1*y4.2 + 1*y4.3

  # equal thresholds
  y1.1 + y1.2 + y1.3 | th11*t1 + th12*t2 + th13*t3
  y2.1 + y2.2 + y2.3 | th21*t1 + th22*t2 + th23*t3
  y3.1 + y3.2 + y3.3 | th31*t1 + th32*t2 + th33*t3
  y4.1 + y4.2 + y4.3 | th41*t1 + th42*t2 + th34*t3

  # between intercepts
  by1 + by2 + by3 + by4 ~ 0*1

  ### Step 5: construct a model at the between level
  # between factor
  fb =~ a*by1 + b*by2 + c*by3 + d*by4

  # residual variances
  # by1 ~~ 0*by1; by2 ~~ 0*by2; by3 ~~ 0*by3; by4 ~~ 0*by4

  # between factors not correlated with the within factors fb ~~ 0*fw1 + 0*fw2 + 0*fw3
'

#fit the model with for example the PML estimation method
Pfit <- sem(model, data = wideData, ordered = paste(rep(c("y1","y2",
  "y3","y4"), 3), rep(1:3, each = 4), sep = "."), estimator = "PML"
  , parameterization = "theta")
summary(Pfit)
```

Appendix E: Summary of the design factors and the model parameters used in the simulation study

Model parameter & design factors	*value*
Var(η_{fw})	4
a	.45
b,c	.3
Var(η_{fb})	1
λ_w, λ_b	.5
Var(z_1), Var(z_2), Var(w)	1
diag($\mathbf{\Theta}_w$)	1
diag($\mathbf{\Theta}_b$)	0
total Var(y)	2.25
mean (z_1, z_1, w)	0
proportions 2-point scale	.50,.50
proportions 4-point scale	.16, .34, .34, .16
Balanced: number of clusters (I_j)	200(3)/1000(3)
Unbalanced: number of clusters (I_j)	90(9),70(6),40(3)/500(9),300(6),200(3)