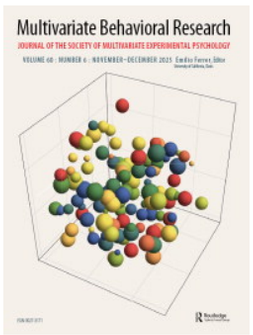
Multiple imputation of missing data in moderated factor analysis

Ginkel, J.R.; Molenaar, D.



Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: www.tandfonline.com/journals/hmbr20

Multiple Imputation of Missing Data in Moderated Factor Analysis

Joost R van Ginkel & Dylan Molenaar

To cite this article: Joost R van Ginkel & Dylan Molenaar (23 Jan 2026): Multiple Imputation of Missing Data in Moderated Factor Analysis, Multivariate Behavioral Research, DOI: 10.1080/00273171.2025.2606868

To link to this article: https://doi.org/10.1080/00273171.2025.2606868

MULTIVARIATE BEHAVIORAL RESEARCH
https://doi.org/10.1080/00273171.2025.2606868

Multiple Imputation of Missing Data in Moderated Factor Analysis

Joost R van Ginkel[a] and Dylan Molenaar[b]

[a]Leiden University; [b]University of Amsterdam

ABSTRACT

In moderated factor analysis, the parameters of the traditional common factor model are a function of an external continuous moderator variable. Handling missing values on the observed indicator variables of the common factors is straightforward as the parameters can be estimated using full information maximum likelihood. However, for cases with missing values on the moderator variable the likelihood function cannot be evaluated. Consequently, in practical applications of the moderated factor model, these cases are omitted from the analysis by listwise deletion. As listwise deletion is known to potentially affect the consistency and precision of the results, we propose a moderated factor model based multiple imputation procedure for handling missing values on the moderator variable in the presence of missing values on the indicator variables. We compare this new procedure with listwise deletion and predictive mean matching. The results show that both listwise deletion and predictive mean matching have less power and produce more bias in parameter estimates than multiple imputation under the moderated factor model.

KEYWORDS

moderated factor analysis; missing data; multiple imputation; full information maximum likelihood; predictive mean matching; listwise deletion

Introduction

Common factor analysis (e.g., Jöreskog, 1969) is a widely used technique for modeling the dimensional structure of tests or questionnaires (e.g., Kline, 2013) like the NEO personality inventory (NEO-PI; McCrae et al., 2005) and the Wechsler's Adult Intelligence scales (WAIS; Wechsler et al., 2008). The underlying model assumes one or more common factors to underlie the observed data which are interpreted in terms of the theoretical constructs that the tests or questionnaires purport to measure (e.g., a personality trait or cognitive ability).

Multigroup extensions of the common factor model have been developed to study differences on the common factors across grouping variables like sex and country (e.g., Jöreskog, 1971). In doing so, the factor model parameters are tested or assumed to be equal across groups, a condition called measurement invariance (Mellenbergh, 1989; Meredith, 1964, 1993; Millsap, 2012). Next, group differences on the common factor means and variances are assumed to reflect group differences on the same underlying psychological construct.

More recently, extensions of the common factor model have been proposed to allow for common factor model parameter differences across continuous background variables like age or SES. In these so-called moderated factor models (Bauer, 2017; Bauer & Hussong, 2009; Hildebrandt et al., 2016; Mehta & Neale, 2005; Molenaar et al., 2010), the common factor model parameters are a mathematical function of the continuous background variable. Similarly, as in multigroup modeling, moderated factor models can be used to test equality of the factor model parameters across the continuous moderator to aid the theoretical interpretation of the results. As such, moderated factor models are a powerful tool to (1) test for measurement invariance across continuous variables (see Bauer, 2017); (2) study common factor mean differences across continuous variables while allowing for common factor heteroscedasticity (which is not straightforward in traditional factor analysis, structural equation modeling, or latent regression); and (3) to test various substantive hypotheses related to differentiation of intelligence (see Breit et al., 2022; Hildebrandt et al., 2016). For example, Hildebrandt et al. (2016) investigated seven broad cognitive ability measures using a moderated one-factor model in the

CONTACT Joost R. Van Ginkel jginkel@fsw.leidenuniv.nl Methodology and Statistics, Leiden University, PO Box 9500, RB Leiden, 2300, The Netherlands.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/00273171.2025.2606868.

light of the age differentiation hypothesis (Breit et al., 2022). The authors found that the loadings of Comprehension-Knowledge and Long-term Retrieval are quadratically related to age (ranging from 5 to 23 years), with Comprehension-Knowledge showing an inverted-U effect while Long-term Retrieval showed a (small but significant) U-effect. In addition, the residual variances of Comprehension-Knowledge, Long-term Retrieval, and Processing Speed were found to be linearly increasing across age, while the residual variance of Fluid Reasoning was found to linearly decrease across age. See Breit et al. for other applications of the moderated factor model to intelligence data. In addition, the moderated factor model has been applied to, for example, positive and negative experience data (Jovanović et al., 2020) and data on delinquent behavior (Bauer, 2017) to test for measurement invariance across age and sex.

Like any other statistical model, estimation of the parameters of the moderated factor model may be complicated by missing values. When the missing data occur on the observed indicators of the common factors, the missing data are relatively unproblematic as parameters of the moderated factor model are generally estimated using full information maximum likelihood. However, if for some respondents, missing data occur on the moderator variable, the likelihood function cannot be evaluated for these respondents.

Consequently, in the software packages that are available to fit moderated factor models, Mplus (Muthén & Muthén, 2017; see the supplementary material of Bauer, 2017, for a tutorial on moderated factor models in Mplus), OpenMx (Boker et al., 2011; see Kolbe et al., 2024 for a tutorial on moderated factor models in OpenMx), SAS (SAS Institute Inc, 2013; see the supplementary material of Bauer & Hussong, 2009 for a tutorial on moderated factor models in SAS), and R package mnlfa (Robitzsch, 2025), cases with missing values on the moderator are automatically deleted listwise (Mplus), or need to be manually deleted listwise to be able to use the software (OpenMx, SAS, and `mnlfa`).[1] Not only does listwise deletion on the basis of the moderator (henceforth denoted by LD-m) lead to a reduction of power, it could also lead to biased results if there are systematic differences between respondents with missing values on the moderator and respondents with observed values on the moderator.

More specifically, for LD-m to give unbiased results, the missing data on the moderator needs to be *missing completely at random* (MCAR; Little & Rubin, 2002). Let $\mathbf{Y}$ be an $N \times J$ matrix of observed indicator scores with an observed part $\mathbf{Y}_{obs}$ and a missing part $\mathbf{Y}_{mis}$, let $\mathbf{m}$ be a N-dimensional vector of observed moderator values with an observed part $\mathbf{m}_{obs}$ and a missing part $\mathbf{m}_{mis}$, and let $\mathbf{r}$ a vector of size N with entry $r_i = 1$ if $m_i \in \mathbf{m}_{obs}$, and $r_i = 0$ if $m_i \in \mathbf{m}_{mis}$ for $i = 1, ., N$. For the specific situation of missing data on $\mathbf{m}$ being MCAR, MCAR is defined as:

$$P(\mathbf{r}|\mathbf{m}_{obs}, \mathbf{m}_{mis}, \mathbf{Y}_{obs}, \mathbf{Y}_{mis}) = P(\mathbf{r}) \quad (1)$$

Thus, if the condition in Equation (1) does not hold, LD-m may bias the moderated factor modeling results. One possible solution is to resort to Bayesian parameter estimation of the moderated factor model (see Enders et al., 2024), where the missing values in m_i are sampled along with the other parameters in the model. However, for researchers who want to rely on a maximum likelihood framework, there is no general solution besides LD-m as of yet. Therefore, in this study, we focus on *multiple imputation* (MI; Rubin, 1987) which allows researchers to 1) account for missing data in their moderated factor analysis; and 2) to keep using the software package that they are familiar with to fit the moderated factor model to data using maximum likelihood (i.e., Mplus, OpenMx, or SAS).

In MI, the missing data are estimated H times using a statistical model that is suitable for the data in question, creating H plausible complete versions of the incomplete dataset. Next, these complete versions are analyzed using the statistical analysis of interest (a moderated factor analysis in our situation), resulting in H outcomes of the same analysis, where the differences are the result of the different estimates for the missing data across the imputed datasets. Finally, the H results are combined into one pooled statistical analysis, using combination rules defined by Rubin (1987).

An advantage of MI over LD-m is not only that all cases are kept in the analysis, but also that the condition in Equation (1) may be loosened to:

$$P(\mathbf{r}|\mathbf{m}_{obs}, \mathbf{m}_{mis}, \mathbf{Y}_{obs}, \mathbf{Y}_{mis}) = P(\mathbf{r}|\mathbf{m}_{obs}, \mathbf{Y}_{obs}) \quad (2)$$

in order for results to be still unbiased, provided that the model that generated the observed values and

[1]We encountered that some researchers are not aware of this. Indeed, many software packages exist (including Mplus and OpenMx, but also for instance the nlsem R package by Umbach et al., 2017 and the approach by Cham et al., 2017) to model interactions in factor models in the presence of missing data. However, these packages all involve interactions among latent variables, while in moderated factor analysis, the latent variable(s) interact with an observed variable, not with a latent variable. This complicates standard full information maximum likelihood as implemented in for example Mplus and OpenMx. That is, the moderator variable needs to be specified as a constraint variable (Mplus) or a definition variable (OpenMx) which cannot contain missing values.

the model used for MI are the same (here, a moderated factor model). When Equation (2) holds, missing data on **m** are said to be missing at random (MAR; Little & Rubin, 2002).

Finally, when the right-hand side of Equation (2) cannot be simplified (i.e., missing data also depend on $\mathbf{m}_{mis}$ and/or $\mathbf{Y}_{mis}$ then the missing data on **m** are said to be *missing not at random* (MNAR; Little & Rubin, 2002). Under MNAR, neither LD-m, nor MI are guaranteed to give unbiased results. However, Schafer (1997, pp. 26–27) pointed out that even under MNAR, MI will give less biased results than listwise deletion does. Consequently, even when there is reason to believe that missing data are MNAR, then still it may be beneficial to prefer MI over LD-m in moderated factor analysis in order to reduce bias.

As already mentioned, in order for MI to give unbiased results under MAR, the imputation model needs to be the same as the model that generated the data. A problem in the context of moderated factor analysis is that currently, no MI procedure to impute under a moderated factor model is available. In standard MI, available in the R package `mice` (R Core Team, 2021; Van Buuren & Groothuis-Oudshoorn, 2011), the underlying statistical model used for imputing the data is either a linear regression model for numerical data (assuming homogeneous parameters across respondents), or multinomial logistic regression for categorical variables. This method is also known as *Fully Conditional Specificati*on (FCI; e.g., Van Buuren, 2012, pp. 108–116; Van Buuren et al., 2006). Additionally, even though Mplus has an option to do multiple imputation under various SEM models, it can only impute under a standard factor model, not under a moderated factor model.

A variant of FCI that is generally assumed to be robust to model violations is called *Predictive mean matching* (PMM; Van Buuren, 2012, p. 68–74; Van Buuren et al., 2006; Van Buuren et al., 1999; Rubin, 1987). In PMM, missing data on categorical variables are handled in the same way as in standard FCS, namely by means of multinomial logistic regression. However, for numerical variables, a different procedure is used.

First, the missing data are predicted using linear regression with other variables as predictors. Next, for each respondent with a missing value on the variable in question, a respondent with an observed value is found with a similar predicted value on this variable according to the regression model. Finally, the respondents with observed values donate their observed values to their matching respondents with missing data. This is done for each variable with missing data, and the whole process is repeated in an iterative manner. For more details, we refer to the references just mentioned.

Because PMM uses actual observed values as imputations rather than simulating imputed values according to the regression model, it generally preserves data structures better when data do not behave according to a linear regression model. The robustness of PMM to model violations has been shown in several simulation studies (e.g., Marshall et al., 2010; Tang et al., 2005; Yu et al., 2007). However, to our knowledge, the robustness of PMM has not been explicitly studied in a common factor analysis context with violations of measurement invariance depending on **m**. This raises the question whether the robustness of PMM to model violations generalizes to a situation where moderated factor analysis is the analysis of interest. It may be possible that PMM is not sufficiently capable of picking up the heterogeneity of parameters across respondents by its matching procedure, and that an imputation method that explicitly imputes under a moderated factor model is necessary for handling the missing data.

In the current study we will propose a multiple imputation approach based on factored specification (FS; Ibrahim, 1990; see also Enders, 2025). That is, we factor the joint distribution of the indicator variables and moderator variables into two distributions: the factor model distribution conditional on the moderator, and a distribution for the moderator. An advantage of this approach is that it works well for models that include non-linear effects (e.g., Lüdtke et al., 2020; Enders, 2025). As such, our approach may be better capable of capturing the heterogeneity in the data due to the moderation effects as compared to PMM. In addition, this procedure can be applied in combination with any of the software packages that are currently available to fit the moderated factor model to data (i.e., Mplus, OpenMx, or SAS).

The outline is as follows: First, we formally introduce the moderated factor model and derive our multiple imputation procedure. Next, we present a simulation study in which we compare the newly proposed method with LD-m and PMM regarding bias and rejection rates of the model parameter estimates. We end with recommendations about which methods should be preferred under different circumstances to handling missing data in moderated factor analysis with missing values on the moderator.

The moderated factor model

We first give an introduction of the common factor model notation that will be used in this paper: Suppose that υ_j is the intercept of variable j ($j = 1, \ldots, J$), λ_{jd} is the factor loading of variable j on factor d ($d = 1, \ldots, D$), $\boldsymbol{\eta}_i$ is a vector of factor scores of person i ($i = 1, ..., N$) with elements η_{id}, and ε_{ij} is a residual. Then the common factor model for observed variable y_{ij} is given by:

$$y_{ij} = \upsilon_j + \sum_{d=1}^{D} \lambda_{jd}\eta_{id} + \varepsilon_{ij} \quad \text{with } \varepsilon_{ij} \sim N(0; \theta_j) \text{ and } \boldsymbol{\eta} \sim MVN(\boldsymbol{\kappa}; \boldsymbol{\Sigma}) \tag{3}$$

where $\boldsymbol{\kappa}$ is a D-dimensional vector containing the factor means and $\boldsymbol{\Sigma}$ is a $J \times J$ factor covariance matrix. In addition, in the above, we assume that a sufficient number of λ_{jd} are fixed to 0 to avoid rotational indeterminacy of the resulting matrix of factor loadings.

An important assumption of the common factor model is that the parameters are the same across all respondents. Random fluctuations of the parameters across respondents can be accounted for in multilevel factor models (e.g., Ansari et al., 2002). However, if the parameters differ systematically across a background variable m_i, the assumption of measurement invariance discussed above is violated which is referred to as *differential item functioning* (DIF; Ackerman, 1992; Mellenbergh, 1989). In the moderated factor model, such parameter differences across m_i can be accommodated by modifying Equation (3) into:

$$y_{ij} = \upsilon_j(m_i) + \sum_{d=1}^{D} \lambda_{jd}(m_i)\eta_{id} + \varepsilon_{ij} \quad \text{with } \varepsilon_{ij} \sim N(0; \theta_j(m_i)) \text{ and } \boldsymbol{\eta}_i \sim MVN(\boldsymbol{\kappa}(m_i); \boldsymbol{\Sigma}(m_i)). \tag{4}$$

That is, the intercepts υ_j, the factor loadings λ_{jd}, the residual variances θ_j, the factor means $\boldsymbol{\kappa}$, and the factor variances $\boldsymbol{\Sigma}$ are made functions of m_i denoted by respectively $\upsilon_j(m_i)$, λ_{jd} (m_i), θ_j (m_i), the factor means $\boldsymbol{\kappa}(m_i)$, and the factor variances $\boldsymbol{\Sigma}(m_i)$. These functions can be of any form in principle, but commonly the model is presented using polynomial functions of order Q. More specifically, for the intercepts, the loadings, the residual variances, and the factor means:

$$\begin{aligned} \upsilon_j(m_i) &= \upsilon_{0j} + \sum_{q=1}^{Q} \upsilon_{qj} m_i^q, \\ \lambda_{jd}(m_i) &= \lambda_{0jd} + \sum_{q=1}^{Q} \lambda_{qjd} m_i^q, \\ \theta_j(m_i) &= \exp\left(\theta_{0j} + \sum_{q=1}^{Q} \theta_{qj} m_i^q\right), \\ \text{and } \kappa_d(m_i) &= \kappa_{0d} + \sum_{q=1}^{Q} \kappa_{qd} m_i^q, \end{aligned} \tag{5}$$

where an exponential function is used for $\theta_j(m_i)$ to prevent negative variances. In the moderated factor model, the diagonal elements of the factor covariance matrix are moderated separately from the off-diagonal elements. That is, if the covariance matrix is given by

$$\boldsymbol{\Sigma}(m_i) = \begin{bmatrix} \sigma_{11}^2(m_i) & & \\ \vdots & \ddots & \\ \sigma_{D1}(m_i) & \cdots & \sigma_{DD}^2(m_i) \end{bmatrix}, \tag{6}$$

the diagonal elements are moderated in a similar way as the residual variances:

$$\sigma_{dd}^2(m_i) = \exp\left(\varphi_{0d} + \sum_{q=1}^{Q} \varphi_{qd} m_i^q\right). \tag{7}$$

To ensure that the resulting moderated covariance matrix $\boldsymbol{\Sigma}(m_i)$ is positive definite across all values of m_i, Bauer (2017) used the following function (based on Fisher's z-transformation) to moderate $\sigma_{dd'}$:

$$\sigma_{dd'}(m_i) = \sqrt{\sigma_{dd}^2 \sigma_{d'd'}^2} \frac{\exp\left[2\rho_{dd'}(m_i)\right] - 1}{\exp\left[2\rho_{dd'}(m_i)\right] + 1}, \left(d \neq d'\right) \tag{8}$$

with $\rho_{dd'}(m_i) = \rho_{0dd'} + \sum_{q=1}^{Q} \rho_{qdd'} m_i^q$.

Note that this ensures a positive definite $\boldsymbol{\Sigma}(m_i)$ for a common factor model with 2 dimensions ($D = 2$). For $D > 2$ this approach does not guarantee positive definiteness of $\boldsymbol{\Sigma}(m_i)$.

In the moderated factor model above, each of the functions $\upsilon_j(m_i)$, λ_{jd} (m_i), θ_j (m_i), $\boldsymbol{\kappa}(m_i)$, $\sigma_{dd}^2(m_i)$, and $\rho_{dd'}(m_i)$ contain baseline parameters which are indexed by 0 (e.g., υ_{0j}, λ_{0j}, etc.) and moderation parameters which are indexed by $1, ..., Q$. The baseline parameters give the parameter values at $m_i = 0$, and the moderation parameters model the shape of the relation between the corresponding common factor model parameters and m_i. For $Q = 1$ the relation between the (log or Fisher z-transformed) parameter and m_i is linear, and for $Q > 1$ the relation becomes increasingly curved. In practice however, applications are mostly limited to linear moderation ($Q = 1$) or, in a few cases, curvilinear moderation ($Q = 2$) in the case of large ranges, for example, an age variable that covers ages from 18 to 70 years. For $Q = 1$, a statistical test on the presence of moderation of a certain parameter in a given dataset involves testing the moderation parameter to be equal to 0 (e.g., for a factor loading one would test $\lambda_{1jd} = 0$).

Estimation of the moderated factor model above is relatively straightforward: The fitting function of the standard common factor model can be adapted to include a predicted covariance matrix, $\boldsymbol{\Sigma}_y|m_i$ and mean vector $\boldsymbol{\mu}_y|m_i$ conditional on m_i instead of $\boldsymbol{\Sigma}_y$

and $\boldsymbol{\mu}_y$ respectively, the predicted covariance and mean vector that are invariant across m_i (see Mehta & Neale, 2005). Optimizing the resulting fitting function will provide maximum likelihood estimates of all baseline and moderation parameters. As the fitting function needs to be evaluated across all data records, this procedure is a full information procedure, meaning that missing values on the dependent variables y_{ij} are relatively unproblematic under MAR. However, as the fitting function is conditional on m_i, for cases with missing values on m_i, the fitting function cannot be evaluated as $\boldsymbol{\Sigma}_y|m_i$ and $\boldsymbol{\mu}_y|m_i$ are undefined. As explained, current implementations of the moderated factor model in Mplus, OpenMx, and SAS involve listwise deletion of cases with missing m_i which is suboptimal. We therefore consider an MI-based approach in the next section.

Multiple imputation for moderated factor analysis

As mentioned above, our MI method for moderated factor models is based on a factored specification. We will therefore refer to it as *Factored Specification—Moderated Factor Analysis* (FS-MFA). Specifically, we factor the joint distribution of the vector of observed indicators $\mathbf{y}_i$ and the moderator variable m_i into a multivariate normal distribution for $\mathbf{y}_i$ conditional on m_i, $f(\mathbf{y}_i|m_i)$, and a standard normal distribution for the moderator, $\varphi(m_i)$:

$$p(\mathbf{y}_i, m_i) = f\left(\mathbf{y}_i|m_i\right) \times \varphi(m_i). \tag{9}$$

The advantage is that we can obtain conditional distributions for m_i and $\mathbf{y}_i$ by using Bayes theorem. With these distributions in place, we can generate values for the missing data on $\mathbf{y}_i$, collected in matrix $\mathbf{Y}_{mis}$, and – the main target – the missing data on m_i collected in vector $\mathbf{m}_{mis}$. By replacing these missing data with different samples from these distributions, multiply imputed datasets are obtained. These datasets can be analyzed using standard maximum likelihood estimation (in e.g., Mplus or OpenMx) to obtain parameter estimates. After pooling these results, the pooled estimates and standard errors can be used to make inferences about the presence of moderation in the data.

As the process of data generation depends on the starting values for $\mathbf{Y}_{mis}$ and $\mathbf{m}_{mis}$, we will use an algorithm known as *Data Augmentation* (DA; Tanner and Wong, 1987). This algorithm consists of a posterior step (P-step) and an imputation step (I-step). The full algorithm is formally presented in the Appendix. Conceptually, the algorithm proceeds as follows:

1. Starting values for $\mathbf{Y}_{mis}$ and $\mathbf{m}_{mis}$ are randomly drawn from a multivariate normal distribution with the observed mean and covariance matrix. These values are filled in for the missing data, resulting in an imputed dataset.
2. The moderated factor model (see below) is fit to the imputed dataset. The parameter estimates and asymptotic covariance matrix of the parameters are recorded.
3. P-step: A new asymptotic covariance matrix is drawn from an inverse Wishart distribution with the estimated asymptotic covariance matrix obtained in Step 1) as scale matrix. Next, new parameter values are drawn from a multivariate normal distribution with the parameter estimates from Step 2) as means and the sampled matrix as the covariance matrix.
4. I-step: Using the sampled parameters from Step 2), data for $\mathbf{m}_{mis}$ are sampled from the distribution of m_i conditional on $\mathbf{y}_i$, $g(m_i|\boldsymbol{y}_i)$, and data for $\mathbf{Y}_{mis}$ are sampled from the distribution of y_{ij} conditional on the other indicators and m_i. The latter is known to be a normal distribution, i.e., $N(y_{ij}|\mathbf{y}_{i,-j}, m_i)$ where $\mathbf{y}_{i,-j}$ denotes the vector of observed indicators without indicator j. The sampled values are filled in for the missing data resulting in a new imputed dataset.

Suppose the data are imputed H times, then steps 1, 2, and 3 are cycled for a prespecified number of $T \times H$ iterations, where at each T'th iteration an imputed dataset is generated using the imputed values from that iteration. The value of T is chosen such that for imputed dataset $h = 1$ $(h = 1, ..., H)$ the imputed values have been sufficiently removed from their starting values at step 0, and that for $h > 1$, the imputed values have been sufficiently removed from the imputed values of imputed dataset $h - 1$. For this study, preliminary simulations showed that $T = 10$ was sufficient for this purpose (based on inspecting convergence plots from a few replicated datasets in the simulation study). However, in more realistic situations where data are not generated using a moderated factor model, 10 iterations may not always be sufficient, and convergence plots may have to provide insight in what an appropriate number of iterations would be.

After obtaining these datasets together with their moderated factor model parameter estimates and standard errors, these results can be pooled using the standard rules by Rubin (1987). Note that in the algorithm, by using an I-step and P-step, both the uncertainty about the missing data and the uncertainty

about the parameters are taken into account. See the Appendix for details.

In this study, we mostly focus on the situation where the imputation model used to create imputed datasets is equivalent to the analysis model that is applied to the imputed datasets for statistical inference. The analysis model can be different from the imputation model, e.g., if moderation effects are dropped one-by-one in a sequence of models. It is, however, recommendable to use an imputation model that is saturated with respect to the moderation effects. That is, all possible moderation effects that are uniquely identified should be included in the model. This will mostly be a model in which all factor loadings, intercepts, and residual variance are moderated, or a model in which an anchor item is identified for which the moderation on the factor loadings and intercepts are dropped, but in which moderation of the factor means and variances is introduced. The latter model is the model we will use in the simulation study below and the real data illustration. According to Schafer (1997, pp. 140–143), using an imputation model that differs from the model used for analysis (e.g., one uses a higher polynomial function than the other, or one includes more moderator variables than the other) is not harmful, as long as the model used for analysis is nested within the model used for imputation.

Simulation study

Fixed design properties

In this simulation study, data were generated under a moderated factor model with one common factor, $N = 500$ subjects, $J = 8$ observed variables, and a standard normal m_i. The model parameters of the moderated factor model are given in Table 1. The choices of these parameters were made such that we could both study type-I error rates for parameters that were 0, and study power for parameters that were not 0. The true values have been roughly based on the real data application described later.

Table 1. Parameters of the simulation model.

j	κ_{01}	κ_{11}	φ_{01}	φ_{11}	υ_{0j}	υ_{1j}	λ_{0j1}	λ_{1j1}	θ_{0j}	θ_{1j}
1	0	0.2	0	0.3	0	0	0.6	0	−1	0
2					0	0	0.6	0.129	−1	0
3					0	0	0.6	0	−1	0.2
4					0	0	0.6	0.129	−1	0.2
5					0	−0.15	0.6	0	−1	0
6					0	−0.15	0.6	0.129	−1	0
7					0	−0.15	0.6	0	−1	0.2
8					0	−0.15	0.6	0.129	−1	0.2

Using the above moderated factor model, 1000 replicated datasets $\mathbf{X} = [\mathbf{Y}, \mathbf{m}]$, were simulated. In these complete datasets, missing data were simulated by removing specific percentages of cells from the data according to several different missingness mechanisms. Finally, the missing data were handled using three different missing-data handling methods. The specific missingness mechanisms, percentages of missing data, and missing-data handling methods will be discussed next.

Independent variables

Missingness mechanism. Missing data were simulated according to MCAR, MAR, and MNAR. Although in the introduction these missingness mechanisms were only defined for missingness on $\mathbf{m}$, it was our intention in our study to simulate missingness on all variables, where the acting missingness mechanism applied for all variables. When MCAR is defined for all variables in the dataset $\mathbf{X}$, the more general definition of MCAR is:

$$P(\mathbf{R}|\mathbf{X}_{obs}, \mathbf{X}_{mis}) = P(\mathbf{R}), \tag{10}$$

where $\mathbf{R}$ is an $N \times (J+1)$ response indicator matrix. MAR is defined as:

$$P(\mathbf{R}|\mathbf{X}_{obs}, \mathbf{X}_{mis}) = P(\mathbf{R}|\mathbf{X}_{obs}), \tag{11}$$

and MNAR is any missingness mechanism where $P(\mathbf{R}|\mathbf{X}_{obs}, \mathbf{X}_{mis})$ cannot be simplified.

In our simulation study, MCAR was simulated as follows. First, an $N \times (J+1)$ matrix $\mathbf{A}$ with uniform random numbers between 0 and 1 was created. Next, for a proportion of missing data p, the $p \times N \times (J+1)$ highest entries in $\mathbf{A}$ were removed in $\mathbf{X}$.

MAR was simulated in the following way. First, like in MCAR, $\mathbf{A}$ was simulated. Next, a latent variable $\mathbf{v}$ of size N with values of $k = 1, 2, 3$ was simulated with all three values having equal probability. Define n_k as the group size of all respondents with $v_i = k$, so that $N = \sum_{k=1}^{K} n_k$. Furthermore, define $\mathbf{m}(k)$ as a subset of $\mathbf{m}$, and $\mathbf{y}_1(k)$ as a subset of variable $\mathbf{y}_1$, for all respondents with $v_i = k$. An $N \times (J+1)$ weight matrix $\mathbf{W}$ was created with the following entries:

1. $w_{ij} = 1$ for $v_i = 1$ and all j,
2. $w_{ij} = m_i\text{-min}[\mathbf{m}(k)] + 0.01$ for $j \leq J$ and $v_i = 2$, and $w_{ij} = 0$, for $j = J+1$ and $v_i = 2$,
3. $w_{ij} = y_{i1}\text{-min}[\mathbf{y}_1(k)] + 0.01$ for $j > 1$ and $v_i = 3$, and $w_{ij} = 0$, for $j = 1$ and $v_i = 3$.

Now suppose that $\mathbf{A}(k)$, $\mathbf{W}(k)$, and $\mathbf{X}(k)$ are the subsets of matrices $\mathbf{A}$, $\mathbf{W}$, and $\mathbf{X}$ respectively, for all

respondents with $v_i = k$. For each k, the $\text{rnd}(p/n_k) \times n_k \times (J+1)$ highest entries of matrix $\mathbf{A}(k) * \mathbf{W}(k)$ were removed in $\mathbf{X}(k)$. If $\sum_{k=1}^{K} \text{rnd}(p/n_k)$ exceeded p because of rounding, then the highest value of $\text{rnd}(p/n_k)$ was rounded downwards to guarantee a proportion of missing data of exactly p.

Conceptually, in the above MAR mechanism, the first group of respondents ($v_i = 1$) has a chance of missing data on both $\mathbf{Y}$ and $\mathbf{m}$ with all entries in $\mathbf{X} = [\mathbf{Y}, \mathbf{m}]$ having equal probability. The second group ($v_i = 2$) has a chance of missing data on $\mathbf{Y}$ but not on $\mathbf{m}$, where the probability of missingness on $\mathbf{Y}$ becomes higher for all variables as $\mathbf{m}$ increases. Finally, the third group has a chance of missing data on $\mathbf{m}$ and variables $\mathbf{y}_2, \ldots, \mathbf{y}_J$, where the probability of missing data on these variables increases as $\mathbf{y}_1$ increases.

Finally, MNAR was simulated in the following way. Define $\mathbf{W} = \boldsymbol{\eta}\mathbf{1}'_{(J+1)} - \left[\min(\boldsymbol{\eta}) + 0.01\right] \times \mathbf{1}_N\mathbf{1}'_{(J+1)}$. The $p \times N \times (J+1)$ highest entries in $\mathbf{A} * \mathbf{W}$ were removed in $\mathbf{X}$. This MNAR mechanism corresponds with a situation where the probability of missing data in $\mathbf{X}$ increases when the unobserved factor score increases.

Proportion of missing data. In total, three different proportions of missing data were simulated: $p = 0.05$, 0.10, 0.30.

Missing-data handling method. Three different methods for handling missing data were studied: LD-m, PMM, and FS-MFA. For methods PMM and FS-MFA, the number of imputations used was $H = 20$. This number was (roughly) based on earlier studies (Van Ginkel, 2020; Van Ginkel et al., 2007). For method PMM, for each variable the missing data were imputed using all the other variables as predictors. The number of iterations used for PMM was $T = 10$ (the default).

The above factors resulted in a full factorial within-subjects design with 3 (missingness mechanism) × 3 (proportion of missing data) × 3 (missing-data handling method) = 18 cells.

Dependent variables

Because the moderated factor model contains many parameters, a selection of parameters was made, to report the results of. These parameters are the ones that are (1) normally the most important for interpretation of the factor model, and (2) expected to be most heavily biased when heterogeneity is not modeled in the imputation model (like in PMM). The studied parameters are λ_{1j1}, θ_{1j}, κ_{11}, and φ_{11}. For the remaining parameters, results are only briefly mentioned.

For each condition, both the bias and rejection rates were studied for these parameters. For parameters that were 0, rejection rates reflect the type-I error rate; for parameters that were not 0, they reflect the power. For methods PMM and FS-MFA, significance tests of the parameters were obtained using Rubin's (1987) combination rules.

Estimation

All three approaches considered involve fitting the true underlying moderated factor model to the (imputed) data. Here we use Mplus (Muthén & Muthén, 2017) in combination with the `MplusAutomation` package in R (Hallquist & Wiley, 2018). The code is available on www.dylanmolenaar.nl, in addition we provide code using OpenMx. This study was not preregistered.

Results

Relationship of missingness with the moderator variable

In the introduction we mentioned that missing data in moderated factor analysis is especially problematic when the missing data on the moderator variable are MNAR. In our simulation study we studied a more general form of MNAR where the probability of missing data on all variables in $\mathbf{X}$ increases when the factor score increases. However, to check whether this MNAR mechanism also implied MNAR for the missingness on the moderator variable specifically, for each replicated dataset we calculated Cohen's d for $\overline{m}_{obs} - \overline{m}_{mis}$. According to Cohen (1988), effects with $0.2 \leq |\text{Cohen's } d| < 0.5$ qualify as small effects, effects with $0.5 \leq |\text{Cohen's } d| < 0.8$ qualify as medium effects, and effects with $|\text{Cohen's } d| \geq 0.8$ are large effects. For each combination of missingness mechanism and percentage of missingness, the average of this Cohen's d is shown in Table 2.

As can be seen from the table, for MNAR and all percentages of missingness, the average Cohen's d qualifies as a small effect. For 5% missing data, the effect even approaches a medium effect. Thus, it is safe to say that missing data on the moderator variable qualify as MNAR.

Results common factor parameters

Bias

For each condition, Cohen's d of the bias was calculated (reference point 0). The results for the bias of κ_{11} and φ_{11} are given in Table 3 (first and third columns). In general, the bias of κ_{11} is (again) smallest for FS-MFA, followed by LD-m. It can be seen that the more the missingness mechanism deviates from MCAR (with MNAR being the most severe deviation from MCAR), the more biased κ_{11} becomes for methods LD-m and PMM.

As the value of p increases, methods LD-m and PMM produce more bias in κ_{11}. Method FS-MFA on the other hand, has small bias across all values of p and all missingness mechanisms. However, the standard deviation of the bias does increase with increasing p.

As for the bias of φ_{11}, only PMM has bias with a discernable effect size. For PMM, this bias increases as the missingness mechanism deviates more from MCAR, and with increasing p. Methods LD-m and FS-MFA do not produce any substantial bias in φ_{11} for any of the missingness mechanisms and any of the values of p.

Table 2. Average Cohen's d for the difference between $\bar{m}_{obs}$ and $\bar{m}_{mis}$ (standard deviations between brackets) for each combination of missingness mechanism and percentage of missingness.

	Percentage of missingness		
	5%	10%	30%
Missingness mechanism	$\bar{d}$ (*SD*)	$\bar{d}$ (*SD*)	$\bar{d}$ (*SD*)
MCAR	0.00 (0.21)	0.00 (0.15)	0.00 (0.10)
MAR	−0.17 (0.26)	−0.11 (0.18)	−0.06 (0.12)
MNAR	*−0.49 (0.20)	*−0.36 (0.14)	*−0.22 (0.12)

*Small effect.

Rejection rate

As a measure of effect size for the rejection rates of the common factor parameters, Cohen's h for differences in proportions was used. In each condition, Cohen's h was calculated for the difference in rejection rate for the specific condition and the rejection rate of the original data. For Cohen's h, the same rules of thumb apply as for Cohen's d.

For both parameters κ_{11} and φ_{11} the power produced by FS-MFA is generally substantially higher than the power produced by methods PMM and LD-m (Table 3, columns 2 and 4). The power produced by PMM is lowest for all values of p and all missingness mechanisms. Furthermore, the power of both parameters produced by FS-MFA is far less influenced by missingness mechanism and the value of p than the power produced by the other two methods.

Table 3. Results for bias (standard deviation between brackets) and rejection rates of the common factor parameters. Entries for bias have been multiplied by 10^3.

			κ_{11}		φ_{11}	
Missingness mechanism	p	Method	M (SD)	Rejection rate	M (SD)	Rejection rate
Original data			−2 (68)	.860	−2 (117)	.731
MCAR	.05	LD-m	−3 (70)	.844	1 (123)	.686
		PMM	−2 (70)	.831	11 (112)	.670
		FS-MFA	−2 (69)	.843	1 (121)	.695
	.10	LD-m	−1 (74)	.796	2 (130)	.655
		PMM	1 (73)	*.783	*22 (108)	*.621
		FS-MFA	1 (73)	.812	−1 (126)	.686
	.30	LD-m	−1 (94)	**.589	8 (172)	**.434
		PMM	2 (88)	**.579	***77 (92)	***.256
		FS-MFA	4 (85)	**.623	−2 (147)	*.500
MAR	.05	LD-m	7 (72)	.801	4 (128)	.674
		PMM	4 (72)	.809	*28 (116)	*.598
		FS-MFA	−1 (71)	.833	0 (125)	.702
	.10	LD-m	8 (74)	*.759	−4 (133)	.649
		PMM	4 (74)	*.765	*33 (112)	*.545
		FS-MFA	−2 (74)	.808	−7 (128)	.690
	.30	LD-m	*16 (85)	**.622	3 (156)	*.532
		PMM	4 (80)	*.695	**59 (105)	**.382
		FS-MFA	−4 (79)	*.764	−6 (139)	*.629
MNAR	.05	LD-m	*29 (74)	*.674	14 (135)	*.577
		PMM	*31 (54)	*.656	*54 (113)	**.445
		FS-MFA	3 (72)	.812	−4 (131)	.654
	.10	LD-m	**40 (78)	**.579	6 (143)	*.539
		PMM	**42 (76)	**.563	**62 (108)	**.369
		FS-MFA	3 (74)	.793	−15 (135)	.649
	.30	LD-m	**51 (94)	***.411	−1 (178)	**.441
		PMM	**53 (83)	***.417	***87 (99)	***.221
		FS-MFA	−6 (83)	*.737	−26 (150)	*.604

*$0.2 \leq$ |Cohen's d/h| < 0.5; ** $0.5 \leq$ |Cohen's d/h| < 0.8; *** |Cohen's d/h| ≥ 0.8.

Table 4. Results for bias (standard deviations between brackets) of the loading parameters. Entries have been multiplied by 10^3.

Missingness mechanism	p	Method	λ_{12} M (SD)	λ_{13} M (SD)	λ_{14} M (SD)	λ_{15} M (SD)	λ_{16} M (SD)	λ_{17} M (SD)	λ_{18} M (SD)
Original data			0 (40)	1 (40)	0 (40)	2 (39)	−1 (43)	0 (41)	0 (40)
MCAR	.05	LD-m	0 (42)	2 (43)	0 (42)	2 (42)	−1 (45)	1 (43)	0 (42)
		PMM	*−12 (38)	2 (38)	*−12 (39)	0 (37)	*−14 (41)	0 (39)	*−13 (38)
		FS-MFA	0 (42)	2 (42)	0 (42)	2 (41)	−1 (44)	0 (43)	0 (41)
	.10	LD-m	1 (45)	1 (45)	0 (45)	2 (44)	0 (48)	1 (46)	0 (45)
		PMM	**−23 (37)	2 (36)	**−23 (37)	0 (35)	**−25 (38)	0 (37)	**−25 (36)
		FS-MFA	0 (45)	1 (43)	−1 (44)	2 (43)	−1 (47)	0 (45)	−1 (44)
	.30	LD-m	2 (64)	4 (63)	2 (64)	3 (63)	0 (66)	2 (66)	0 (63)
		PMM	***−60 (33)	5 (31)	***−57 (33)	0 (30)	***−64 (33)	1 (32)	***−62 (33)
		FS-MFA	−3 (59)	2 (57)	−1 (57)	0 (56)	−5 (59)	1 (58)	−4 (56)
MAR	.05	LD-m	−1 (44)	0 (44)	−2 (45)	3 (44)	−2 (46)	0 (46)	−1 (44)
		PMM	*−12 (41)	6 (39)	*−11 (41)	7 (39)	*−13 (42)	5 (41)	*−11 (40)
		FS-MFA	−1 (44)	0 (44)	−1 (44)	2 (44)	−2 (46)	−1 (45)	−1 (43)
	.10	LD-m	−3 (46)	−2 (47)	−2 (46)	1 (45)	−4 (50)	−2 (47)	−3 (45)
		PMM	**−22 (39)	5 (38)	*−19 (39)	5 (36)	**−24 (41)	4 (38)	**−21 (38)
		FS-MFA	−3 (45)	−2 (46)	−3 (45)	0 (44)	−4 (48)	−3 (46)	−3 (44)
	.30	LD-m	0 (57)	0 (58)	1 (57)	3 (58)	−2 (59)	1 (58)	−1 (57)
		PMM	***−39 (37)	6 (34)	***−38 (37)	5 (34)	***−43 (38)	6 (35)	***−39 (37)
		FS-MFA	−4 (50)	−3 (51)	−3 (49)	−1 (50)	−6 (53)	−1 (51)	−4 (50)
MNAR	.05	LD-m	−8 (46)	1 (46)	−8 (47)	0 (44)	−9 (48)	0 (47)	−7 (45)
		PMM	**−26 (40)	4 (39)	**−24 (41)	5 (37)	**−27 (41)	4 (40)	**−25 (40)
		FS-MFA	−6 (45)	1 (46)	−5 (46)	0 (45)	−6 (48)	−1 (47)	−3 (45)
	.10	LD-m	*−10 (50)	1 (49)	−9 (50)	1 (48)	*−11 (51)	0 (51)	−9 (48)
		PMM	***−38 (39)	5 (37)	***−35 (39)	6 (36)	***−38 (40)	6 (38)	***−37 (38)
		FS-MFA	−8 (49)	1 (48)	−5 (49)	0 (47)	−9 (50)	0 (50)	−5 (48)
	.30	LD-m	−8 (65)	2 (63)	−6 (63)	4 (62)	−8 (64)	2 (64)	−5 (61)
		PMM	***−53 (36)	6 (35)	***−51 (38)	8 (34)	***−53 (37)	9 (35)	***−50 (35)
		FS-MFA	−7 (56)	0 (54)	−7 (55)	0 (55)	−10 (57)	−1 (56)	−4 (54)

*$0.2 \leq$ |Cohen's d| < 0.5; ** $0.5 \leq$ |Cohen's d| < 0.8; *** |Cohen's d| ≥ 0.8.

Results loading parameters

Bias

Like the results of the common factor parameters, Cohen's d was used as a measure of effect size for the bias of the loading parameters. Table 4 displays the bias of the loading parameters, along with the effect sizes. For comparison, the results of the original data without missing data are displayed as well.

The table shows that in general, PMM has a large bias in the loading parameters in all conditions. This bias increases with increasing p. Methods LD-m and FS-MFA have relatively small bias, with FS-MFA usually having a slightly smaller bias than LD-m. However, for FS-MFA the standard deviation of the bias is generally smaller than for LD-m, which suggests more stable estimates of the loading parameters. Missingness mechanism seems to have a relatively small impact on the bias for all methods.

Rejection rate

For the rejection rates of the loading parameters, again Cohen's h for proportions was used as a measure of effect size. The results for rejection rates are displayed in Table 5.

Like the results for bias of the loading parameters, the results for PMM are substantially worse than for the other two methods. For the nonzero parameters ($\lambda_{12}, \lambda_{14}, \lambda_{16}, \lambda_{18}$), the power of PMM is much lower in terms of effect size than that of LD-m and FS-MFA, in all conditions. Overall, method FS-MFA has the highest power, although it differs little from LD-m. For the zero parameters ($\lambda_{13}, \lambda_{15}, \lambda_{17}$), the type-I error rates are also much lower than those of methods LD-m and FS-MFA. For the latter two methods, the type-I error rate never differs from the type-I error rate of the original data with a discernible effect size.

When p increases, the power drops for all methods. However, for methods LD-m and FS-MFA, the type-I error rate remains relatively constant across different values of p. This is in accordance with what one would expect. As for the effect of missingness mechanism, the power of PMM is most affected by missingness mechanisms that deviate from MCAR (MAR and MNAR), followed by LD-m. With the exception of method PMM, type-I error rates are not heavily influenced by the different missingness mechanisms.

On a side note, with 30% missing data, the power is relatively low for all methods. Even for FS-MFA the power lies somewhere between 0.50 and 0.60. However, this does not invalidate FS-MFA, but this is

Table 5. Results for rejection rates of the loading parameters.

			Rejection rate						
Missingness mechanism	p	Method	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}
Original data			.833	.074	.844	.075	.816	.067	.848
MCAR	.05	LD-m	.807	.067	.798	.065	.777	.070	.811
		PMM	*.724	.041	*.709	.041	*.693	.042	*.716
		FS-MFA	.806	.061	.804	.066	.776	.068	.812
	.10	LD-m	.762	.068	*.754	.074	*.729	.069	*.744
		PMM	**.568	*.027	**.580	*.019	**.551	*.023	**.526
		FS-MFA	.774	.058	*.754	.065	.749	.070	*.750
	.30	LD-m	**.506	.069	**.536	.085	**.519	.084	**.512
		PMM	***.085	**.000	***.077	*.001	***.053	*.003	***.057
		FS-MFA	**.507	.051	**.546	.066	**.508	.055	**.516
MAR	.05	LD-m	.771	.080	.770	.068	.767	.083	.784
		PMM	*.678	.042	*.705	.038	*.664	.050	*.700
		FS-MFA	.775	.064	.772	.064	.762	.078	.779
	.10	LD-m	*.689	.072	*.723	.060	*.688	.077	*.724
		PMM	**.545	*.029	**.588	*.024	**.531	.027	**.553
		FS-MFA	*.720	.058	*.728	.055	*.703	.067	*.724
	.30	LD-m	**.578	.079	**.586	.063	**.563	.067	**.591
		PMM	***.285	.008	***.292	*.009	.275	*.008	***.302
		FS-MFA	*.633	.061	*.635	.066	*.621	.070	**.633
MNAR	.05	LD-m	*.697	.072	*.686	.069	*.674	.080	*.689
		PMM	**.482	*.023	***.468	*.024	**.472	.031	**.487
		FS-MFA	*.706	.064	*.709	.066	*.695	.074	*.733
	.10	LD-m	*.619	.059	**.627	.075	*.597	.072	**.631
		PMM	***.311	*.010	***.317	.008	***.320	*.009	***.333
		FS-MFA	*.631	.057	*.670	.066	*.611	.064	*.670
	.30	LD-m	**.489	.075	**.491	.081	**.487	.074	**.515
		PMM	***.116	*.005	***.124	*.003	***.129	*.002	***.134
		FS-MFA	**.529	.058	**.552	.070	**.520	.065	**.562

*$0.2 \leq$ |Cohen's h| < 0.5; ** $0.5 \leq$ |Cohen's h| < 0.8; *** |Cohen's h| ≥ 0.8.

merely caused by the fact that 30% missing data results in an underpowered situation anyway. Even under 30% missing data, the type-I error rates and the bias of FS-MFA are still acceptable, showing that FS-MFA still produces valid results in a situation with low power.

Results residual variances

Bias

The results for bias in the residual variance parameters are displayed in Table 6. Bias of the original data is displayed for comparison as well.

It can be seen from the table that the zero parameters ($\theta_{11}, \theta_{12}, \theta_{15}, \theta_{16}$) usually have bias close to 0 for all methods, with only PMM occasionally having a bias with a small effect size (for example, $p = 0.30$, MCAR). The same is not true for the non-zero parameters ($\theta_{13}, \theta_{14}, \theta_{17}, \theta_{18}$), where PMM produces biases with small to large effect sizes. LD-m and FS-MFA generally produce biases of non-zero parameters close to 0.

As p increases, the bias of the non-zero parameters increases for PMM. For the other methods the bias remains stable across different values of p. However, the standard deviation of the bias increases with increasing p. For FS-MFA this increase is smallest, which again suggests that FS-MFA produces more stable estimates than LD-m does. In general, missingness seems to have little effect on the bias of the residual variance parameters.

More generally, the effect of MNAR on the bias of the loading parameters and residuals is less clearly visible than for the bias of the common factor parameters (Table 3). This makes sense as MNAR was simulated such that the missingness depended directly on the values of $\boldsymbol{\eta}$. If higher scores of $\boldsymbol{\eta}$ have more missing values, then consequently, the mean κ_{11} of $\boldsymbol{\eta}$ will be biased as well. The loading and residual parameters, on the other hand, are affected by MNAR less severely because there is no direct relation between the missing data and these parameters.

Rejection rate

The results for rejection rates of the residual variance parameters, along with the results of the effect sizes, can be found in Table 7.

The table shows that for the zero parameters, type-I error rates of methods LD-m and FS-MFA differ little from the type-I error rates of the original data, while for PMM the type-I error rate is generally lower than that of the original data with small effect sizes. As for power, PMM has a substantially lower power than the other two methods for all non-zero parameters for all p and missingness mechanisms.

Table 6. Results for bias (standard deviations between brackets) of the residual parameters. Entries have been multiplied by 10^3.

			θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{16}	θ_{17}	θ_{18}
Missingness mechanism	*p*	Method	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Original data			5 (75)	0 (78)	2 (73)	0 (73)	0 (74)	−2 (73)	1 (74)	−2 (73)
MCAR	.05	LD-m	6 (79)	−1 (83)	2 (79)	0 (76)	−1 (79)	−3 (78)	0 (79)	−1 (78)
		PMM	5 (71)	2 (75)	*−18 (70)	*−19 (67)	1 (71)	0 (69)	*−20 (71)	*−20 (70)
		FS-MFA	6 (79)	0 (83)	1 (78)	0 (76)	0 (80)	−3 (77)	0 (78)	−1 (77)
	.10	LD-m	6 (86)	−2 (86)	4 (84)	3 (83)	−3 (86)	−3 (83)	0 (87)	−4 (83)
		PMM	6 (68)	4 (69)	**−36 (66)	**−35 (66)	1 (68)	1 (66)	**−40 (69)	**−42 (67)
		FS-MFA	6 (84)	−1 (86)	3 (83)	1 (83)	−2 (84)	−4 (82)	−1 (85)	−5 (83)
	.30	LD-m	12 (120)	−2 (112)	6 (117)	−4 (116)	0 (119)	−1 (116)	1 (120)	−3 (117)
		PMM	8 (54)	*15 (57)	***−101 (53)	***96 (54)	5 (54)	*12 (53)	***−104 (54)	***−99 (54)
		FS-MFA	11 (111)	−3 (117)	0 (109)	−7 (108)	−2 (113)	−2 (110)	−4 (113)	−6 (110)
MAR	.05	LD-m	4 (79)	−3 (83)	2 (79)	0 (79)	−1 (78)	−2 (78)	0 (81)	−2 (78)
		PMM	11 (70)	5 (73)	*−20 (71)	*−16 (71)	0 (69)	6 (68)	*−23 (71)	*−18 (70)
		FS-MFA	3 (79)	−3 (83)	2 (79)	−1 (79)	−2 (78)	−3 (78)	−1 (81)	−3 (78)
	.10	LD-m	5 (85)	−1 (89)	1 (85)	0 (83)	0 (85)	−1 (84)	−2 (86)	−1 (84)
		PMM	12 (69)	10 (71)	**−35 (67)	**−28 (67)	2 (67)	10 (67)	**−39 (69)	**−33 (66)
		FS-MFA	1 (84)	−2 (88)	0 (83)	−1 (81)	−1 (83)	−2 (83)	−3 (84)	−3 (82)
	.30	LD-m	5 (104)	1 (111)	−1 (107)	−1 (106)	−3 (107)	−5 (107)	2 (110)	−1 (107)
		PMM	11 (63)	*15 (64)	***−68 (63)	***−58 (63)	2 (61)	14 (61)	***−70 (64)	***−62 (64)
		FS-MFA	0 (94)	−2 (99)	−2 (93)	−3 (93)	−5 (95)	−6 (96)	−4 (100)	−5 (96)
MNAR	.05	LD-m	5 (80)	0 (82)	1 (78)	1 (79)	0 (79)	−1 (79)	1 (78)	0 (79)
		PMM	6 (72)	6 (74)	*−17 (71)	*−15 (71)	0 (71)	5 (70)	*−20 (70)	*−16 (71)
		FS-MFA	3 (80)	−1 (82)	0 (77)	0 (78)	−1 (79)	−2 (78)	0 (77)	−2 (78)
	.10	LD-m	5 (83)	−2 (87)	0 (85)	−2 (83)	0 (83)	−1 (83)	0 (82)	0 (84)
		PMM	6 (68)	6 (71)	*−34 (69)	*−32 (68)	1 (66)	8 (66)	**−37 (67)	*−32 (69)
		FS-MFA	3 (82)	−3 (86)	−2 (83)	−3 (82)	−1 (82)	−3 (82)	−1 (81)	−1 (83)
	.30	LD-m	7 (107)	−1 (108)	1 (107)	2 (104)	−4 (105)	−2 (104)	1 (106)	2 (104)
		PMM	7 (61)	8 (63)	***−70 (63)	***−65 (61)	0 (61)	9 (60)	***−73 (61)	***−67 (62)
		FS-MFA	3 (95)	−5 (99)	−4 (97)	0 (92)	−4 (98)	−3 (96)	0 (96)	−5 (95)

*$0.2 \leq$ |Cohen's *d*| < 0.5; ** $0.5 \leq$ |Cohen's *d*| < 0.8; *** |Cohen's *d*| ≥ 0.8.

Table 7. Results for rejection rates of the residual parameters.

			Rejection rate							
Missingness mechanism	*p*	Method	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{16}	θ_{17}	θ_{18}
Original data			.055	.070	.784	.783	.054	.051	.775	.806
MCAR	.05	LD-m	.063	.083	.749	.746	.062	.066	.727	.759
		PMM	.039	.048	*.661	*.673	.033	.039	*.652	*.687
		FS-MFA	.058	.079	.747	.741	.060	.056	.727	.762
	.10	LD-m	.056	.069	.703	.694	.064	.056	*.651	*.682
		PMM	*.018	*.024	**.533	**.508	.026	*.016	**.498	**.494
		FS-MFA	.054	.069	*.695	.696	.059	.056	*.657	*.683
	.30	LD-m	.076	.069	**.440	**.434	.068	.064	**.425	**.447
		PMM	*.001	**.000	***.059	***.069	*.001	*.001	***.050	***.053
		FS-MFA	.055	.058	**.414	**.427	.054	.052	**.420	***.421
MAR	.05	LD-m	.057	.070	.722	.742	.057	.058	.716	.740
		PMM	.040	.037	*.649	*.675	.028	.030	*.637	*.664
		FS-MFA	.059	.067	.723	.739	.047	.054	.717	.728
	.10	LD-m	.067	.064	*.678	.683	.060	.063	*.660	*.702
		PMM	.026	*.027	**.518	*.566	.022	.023	**.510	**.553
		FS-MFA	.061	.061	*.690	*.676	.044	.060	*.657	*.679
	.30	LD-m	.069	.081	**.493	**.500	.063	.066	**.504	**.527
		PMM	*.010	*.011	***.255	***.290	*.007	*.009	***.249	***.281
		FS-MFA	.056	.061	**.544	*.557	.047	.048	*.547	**.564
MNAR	.05	LD-m	.064	.080	.729	.745	.056	.058	.730	.754
		PMM	.037	.045	*.675	*.657	.035	.029	*.650	*.681
		FS-MFA	.064	.078	.729	.740	.055	.052	.742	.755
	.10	LD-m	.062	.063	*.682	*.686	.056	.067	*.671	*.712
		PMM	.025	.028	**.542	**.534	.019	*.014	**.510	**.563
		FS-MFA	.058	.061	*.686	*.679	.048	.065	*.680	*.703
	.30	LD-m	.067	.082	**.534	*.555	.067	.059	**.529	**.564
		PMM	*.004	*.005	***.244	***.245	*.005	*.000	***.231	***.242
		FS-MFA	.046	.050	*.571	*.587	.058	.061	*.564	*.580

*$0.2 \leq$ |Cohen's *h*| < 0.5; ** $0.5 \leq$ |Cohen's *h*| < 0.8; *** |Cohen's *h*| ≥ 0.8.

Furthermore, as p increased, the power dropped for the non-zero parameters for all methods. However, for all values of p, the power was highest for FS-MFA. Finally, the effect of missingness mechanism is difficult to interpret as it differs across different missingness mechanisms, different values of p, and different methods. For example, for θ_{14} and $p = 0.05$, (fourth column), LD-m has the highest power under MCAR (0.746), followed by MNAR (0.745) and MAR (0.742). However, for θ_{14} and $p = 0.30$, FS-MFA has the highest power under MNAR (0.587), followed by MAR (0.557) and MCAR (0.427).

Other parameters

For the remaining parameters, the results of methods LD-m, PMM, and FS-MFA are more similar. For the θ_{0j} parameters, PMM produces even less biased results than the other methods and the original data. Furthermore, in general, the remaining parameters are less biased and have rejection rates more similar to those of the original data, than the parameters discussed earlier.

Application

Below we apply the LD-m, PMM, and FS-MFA approaches to a dataset on cognitive ability with severe missingness in the SES moderator.

Data

We analyze the dataset pertaining to cognitive ability published in (Osborne, 1980). The data contains various cognitive ability measures for 477 twins between the ages of 12 and 20. The dataset also contains a SES variable which is an aggregated measure of level of occupation (scale 1–7) and level of education (scale 1–7) of both the father and mother of the twins.

In the current analyses we selected 8 cognitive ability measures which seem to measure a cognitive ability related to perceptual organization: the cube comparison test, the simple arithmetic test, the surface development test, the form board test, the paper folding test, the object aperture test, the Newcastle spatial test, and the mazes test. We used the data of the first member of each twin pair. The resulting data fit well to a one-factor model (RMSEA: 0.042; CFI: 0.991; TLI: 0.998) according to the guidelines by (Schermelleh-Engel et al., 2003). See, Table 8 for the correlation matrix of the cognitive ability measures and the moderator (SES).

In the one-factor model we introduce moderation of all factor model parameters by the SES variable. The first indicator (cube comparison) is assumed to be unmoderated across SES for reasons of identification (anchor variable). As a result, DIF can be investigated by testing the moderation parameters to differ significantly from 0. If the moderation effects are limited to the factor mean and/or variance, it can be concluded that the scale complies to measurement invariance.

Table 8. Correlation matrix of the observed variables in the application.

	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}	y_{i6}	y_{i7}	y_{i8}	m_i
y_{i1}	1								
y_{i2}	.458	1							
y_{i3}	.590	.532	1						
y_{i4}	.494	.462	.636	1					
y_{i5}	.534	.475	.627	.630	1				
y_{i6}	.480	.343	.522	.470	.563	1			
y_{i7}	.606	.611	.720	.690	.705	.568	1		
y_{i8}	.330	.305	.353	.365	.362	.262	.423	1	
m_i	.296	.406	.375	.371	.330	.258	.460	.303	1

For the SES variable, 61% of the cases have a missing value (291 cases). For the observed indicators 444 cases have a full data record, 31 cases of one missing value, and 2 cases have 2 missing values. We standardized all variables (the indicator variables and the moderator). We apply the LD-m, PMM, and FS-MFA approaches using the same procedure as in the simulation study. To correct for multiple testing, we used $\alpha = 0.01$ (two sided), rather than the $\alpha = 0.05$ from the simulation study. Note that for a simulation study, capitalization of chance is not an issue; what is important is that the empirical type-I error rate closely resembles the theoretical type-I error rate. The Rcode is available from www.dylanmolenaar.nl.

Results

See Table 9 for the parameter estimates and standard errors as obtained using the three missing data handling approaches. As can be seen by the results for κ_1, all three approaches show that the common factor mean increases significantly across SES. However, LD-m seems to underestimate the effect substantially as compared to the estimate of PMM and FS-MFA. In addition, FS-MFA identifies two factor loadings that differ across SES: the simple arithmetic loading decreases across SES, while the surface development loading increases across SES. Thus, these two subtests violate measurement invariance across SES. In addition, FS-MFA identifies three residual variances that are moderated by SES. The two other approaches fail to identify both the moderation effects on the loadings and on the residual variances, which illustrates our results from the simulation study and emphasizes the

Table 9. Parameter estimates (standard errors) of the moderation parameters in the application.

Parameter	j	LD-m	PMM	FS-MFA
υ_{1j}	2	0.169 (0.093)	0.078 (0.086)	0.028 (0.068)
	3	0.107 (0.073)	0.038 (0.082)	0.101 (0.066)
	4	0.091 (0.085)	−0.018 (0.087)	0.084 (0.064)
	5	0.000 (0.093)	−0.080 (0.071)	−0.094 (0.083)
	6	0.015 (0.076)	−0.029 (0.086)	0.013 (0.065)
	7	0.094 (0.087)	0.045 (0.080)	−0.007 (0.072)
	8	0.150 (0.087)	0.140 (0.080)	0.162 (0.067)
λ_{1j}	2	−0.088 (0.063)	−0.048 (0.044)	**−0.103 (0.033)**
	3	0.137 (0.069)	0.118 (0.046)	**0.140 (0.036)**
	4	0.117 (0.073)	0.043 (0.052)	0.070 (0.034)
	5	0.042 (0.066)	0.023 (0.044)	0.033 (0.035)
	6	0.060 (0.057)	0.057 (0.052)	0.092 (0.046)
	7	0.019 (0.075)	0.020 (0.044)	0.013 (0.037)
	8	−0.025 (0.102)	−0.003 (0.056)	−0.023 (0.046)
θ_{1j}	1	0.264 (0.103)	0.224 (0.102)	**0.408 (0.085)**
	2	0.172 (0.123)	0.078 (0.080)	0.143 (0.085)
	3	0.053 (0.117)	0.086 (0.087)	0.135 (0.092)
	4	0.253 (0.161)	0.219 (0.121)	**0.404 (0.109)**
	5	−0.172 (0.124)	−0.068 (0.109)	−0.161 (0.116)
	6	0.026 (0.093)	0.177 (0.095)	**0.310 (0.085)**
	7	0.210 (0.151)	0.052 (0.143)	0.334 (0.154)
	8	−0.080 (0.094)	0.035 (0.082)	0.061 (0.080)
κ_1		**0.394 (0.118)**	**0.502 (0.098)**	**0.592 (0.097)**
σ^2_{11}		**0.266 (0.190)**	**0.163 (0.132)**	**0.273 (0.104)**

Note. Parameter estimates that are significant at a 0.01 significance level (two sided) are in boldface. In addition, 1: the cube comparison test; 2: the simple arithmetic test; 3: the surface development test; 4: the form board test; 5: the paper folding test; 6: the object aperture test; 7: the Newcastle spatial test; 8: the mazes test.

importance of accounting for missing data in the moderator of a moderated factor analysis.

Discussion

In this study, the MI procedure FS-MFA was proposed to account for missing data on a continuous moderator variable as an alternative to listwise deletion (LD-m). In a simulation study, both the bias and rejection rate of the parameters of the moderated factor model produced by FS-MFA was compared with the bias and rejection rate produced by two other missing-data methods, namely multiple imputation using predictive mean matching (PMM) and LD-m.

FS-MFA produced less bias and had higher power than the other two methods for almost all parameters, and in almost all situations. PMM generally produced the largest bias and had by far the lowest power of all three methods. Additionally, unlike the other two methods, PMM also produced deflated type-I error rates for the zero parameters. This is probably due to PMM being sensitive to the violations of measurement invariance that was incorporated in our simulation design. FS-MFA takes these violations into account in the model.

Although FS-MFA outperformed both LD-m and PMM regarding bias and rejection rates, the differences between FS-MFA and LD-m were small for most model parameters. However, regarding the mean and variance of the factor scores, FS-MFA did show a clear advantage over LD-m, especially for missingness mechanisms that deviated from MCAR. This is an important result as these parameters are commonly of key interest in moderated factor modeling.

Besides the results discussed in the results section, some additional results were requested by one of the reviewers. To save space, these results are included in supplementary material. The supplementary material firstly includes the results of relative bias and the root mean squared error (RMSE). The results of these two outcome measures did not alter the conclusions about FS-MFA. However, PMM had a lower RMSE on average than the other methods, while both its bias and relative bias are substantial. This implies that method PMM produces substantially biased parameters in moderated factor analysis but does this in a very consistent way.

Secondly, the supplemental material includes simulation results based on the empirical data example by Osborne (1980). The reviewer noted that the circumstances of the empirical data example differed quite from the circumstances in our simulation study, especially with regards to the amount of missing data on the moderator variable (61% in the data example versus 30% in the simulation study at most). Using the exact parameter estimates of the dataset from Osborne (1980) (missing data handled using FS-MFA) as input for the simulation model, incomplete data were simulated of the same sample size, and with the same missingness pattern as the data from Osborne. Although in these simulations the power was generally lower than in the simulation study discussed in the results section, FS-MFA still showed the highest power of all methods, and differences in power between FS-MFA and LD-m were on average somewhat larger than in the simulation study discussed in the results section.

Finally, we will discuss a few shortcomings of the current research. The goal of the current study was to study three principles, namely that 1) for missing data in moderated factor analysis, listwise deletion may lead to biased results under missingness mechanisms deviating from MCAR, 2) multiple imputation using predictive mean matching may not be an optimal approach to account for missingness in moderated factor analysis, and 3) explicitly imputing the missing values according to the moderated factor model is a better approach to handle missing values on the moderator. In this study, we showed each of these three principles. However, firstly, to keep the number of parameters to report manageable, we only looked at a moderated one-factor model where the parameters were linearly related to **m**. Since the results of the

current study consistently pointed toward FS-MFA as the preferred missing-data handling method in moderated factor analysis, we expect that this will also be found in situations with, for example, two factors and quadratic relations of **m** with the model parameters.

Secondly, the FS-MFA method is not based on FCS, but on an algorithm that more closely resembles a procedure known as *joint modeling* (e.g., Schafer, 1997; Van Buuren, 2012, pp. 105–108). A disadvantage of joint modeling as compared to FCS, is that the procedure is less flexible regarding the inclusion of variables for predicting the missing data on other variables. Since one joint model is used for imputation, a variable is either included in the joint model or it is not. Even though strictly speaking, FS-MFA is not exactly a joint modeling procedure, it does share this disadvantage with joint modeling.

In FCS, however, missing data on a specific variable may be predicted by a selection of variables that correlate highly with this variable, while missing data on another variable may be predicted by other variables. This feature of FCS is especially useful when there are many variables in the data, and there is a risk of overfitting when one joint model is used to model the missing data on all variables with missing data. However, developing a FCS version of multiple imputation under a moderated factor model may not be trivial, especially not in situations where **m** is related to the model parameters with higher polynomial functions. This could be a topic of future research, so that a more flexible version of MI that can handle DIF becomes available.

Finally, we only considered the moderated factor model with continuous indicator variables. In the case of categorical indicator variables (i.e., items), moderation of the factor model parameters can be studied in a moderated item response theory model or moderated item factor model (also referred to as moderated non-linear factor analysis). However, the present multiple imputation procedure cannot readily be applied to these situations as the distribution of the items cannot be safely assumed to be normal. Future research can address this issue by developing multiple imputation procedures for the moderated item factor model.

To summarize, FS-MFA is successful at producing unbiased or only slightly biased parameter estimates in moderated factor analysis with missing data. In addition, it preserves the power of the parameter estimates better than LD-m and PMM, even under MAR or MNAR mechanisms. Thus, it may be a good alternative to LD-m when handling missing data in moderated factor analysis.

Article information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was not supported by a grant.

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions is not intended and should not be inferred.

References

Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, *29*(1), 67–91. https://doi.org/10.1111/j.1745-3984.1992.tb00368.x

Ansari, A., Jedidi, K., & Dube, L. (2002). Heterogeneous factor models: A Bayesian approach. *Psychometrika*, *67*(1), 49–77. https://doi.org/10.1007/BF02294709

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*(3), 507–526. https://doi.org/10.1037/met0000077

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*(2), 101–125. https://doi.org/10.1037/a0015583

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P., & Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, *76*(2), 306–317. https://doi.org/10.1007/s11336-010-9200-6

Breit, M., Brunner, M., Molenaar, D., & Preckel, F. (2022). Differentiation hypotheses of intelligence: A systematic review of the empirical evidence and an agenda for future research. *Psychological Bulletin*, *148*(7–8), 518–554. https://doi.org/10.1037/bul0000379

Cham, H., Reshetnyak, E., Rosenfeld, B., & Breitbart, W. (2017). Full information maximum likelihood estimation for latent variable interactions with incomplete indicators. *Multivariate Behavioral Research*, *52*(1), 12–30. https://doi.org/10.1080/00273171.2016.1245600

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Enders, C. K. (2025). Missing data: An update on the state of the art. *Psychological Methods*, *30*(2), 322–339. https://doi.org/10.1037/met0000563

Enders, C. K., Vera, J. D., Keller, B. T., Lenartowicz, A., & Loo, S. K. (2024). Building a simpler moderated nonlinear factor analysis model with Markov chain Monte Carlo estimation. *Psychological Methods*. https://doi.org/10.1037/met0000712

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 621–638. https://doi.org/10.1080/10705511.2017.1402334

Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research*, *51*(2–3), 257–258. https://doi.org/10.1080/00273171.2016.1142856

Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association*, *85*(411), 765–769. https://doi.org/10.1080/01621459.1990.10474938

Jovanović, V., Lazić, M., Gavrilov-Jerković, V., & Molenaar, D. (2020). The scale of positive and negative experience (SPANE). *European Journal of Psychological Assessment*, *36*(4), 694–704. https://doi.org/10.1027/1015-5759/a000540

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183–202. https://doi.org/10.1007/BF02289343

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426. https://doi.org/10.1007/BF02291366

Kline, P. (2013). *Handbook of psychological testing*. Routledge.

Kolbe, L., Molenaar, D., Jak, S., & Jorgensen, T. D. (2024). Assessing measurement invariance with moderated nonlinear factor analysis using the R package OpenMx. *Psychological Methods*, *29*(2), 388–406. https://doi.org/10.1037/met0000501

Lüdtke, O., Robitzsch, A., & West, S. G. (2020). Regression models involving nonlinear effects with missing data: A sequential modeling approach using Bayesian estimation. *Psychological Methods*, *25*(2), 157–181. https://doi.org/10.1037/met0000233

Marshall, A., Altman, D. G., Royston, P., & Holder, R. L. (2010). Comparison of techniques for handling missing covariate data with prognostic modelling studies: A simulation study. *BMC Medical Research Methodology*, *10*, 7. https://doi.org/10.1186/1471-2288-10-7

McCrae, R. R., Costa, P. T., Jr, & Martin, T. A. (2005). The NEO–PI–3: A more readable revised NEO personality inventory. *Journal of Personality Assessment*, *84*(3), 261–270. https://doi.org/10.1207/s15327752jpa8403_05

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*(3), 259–284. https://doi.org/10.1037/1082-989X.10.3.259

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, *13*(2), 127–143. https://doi.org/10.1016/0883-0355(89)90002-5

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, *29*(2), 177–185. https://doi.org/10.1007/BF02289699

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. J. (2010). Modeling differentiation of cognitive abilities within the higher order factor model using moderated factor analysis. *Intelligence*, *38*(6), 611–624. https://doi.org/10.1016/j.intell.2010.09.002

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Osborne, R. T. (1980). *Twins: black and white*. Foundationfor Human Understanding.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Wiley.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Robitzsch, A. (2025). mnlfa: Moderated nonlinear factor analysis. R package version 0.4-35 [Computer code]. Retrieved from https://CRAN.R-project.org/package=mnlfa

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.

SAS Institute Inc. (2013). *SAS/STAT software, version 13.2*. http://www.sas.com/

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significanceand descriptive goodness-of-fit measures. *Methods of Psychologicalresearch Online*, *8*(2), 23–74. http://www.mpr-online.de

Tang, L., Song, J., Belin, T. R., & Unützer, J. (2005). A comparison of imputation methods in a longitudinal randomized clinical trial. *Statistics in Medicine*, *24*(14), 2111–2128. https://doi.org/10.1002/sim.2099

Tanner, M. A., & Wong, W. H. (1987). The calculation ofposterior distributions by data augmentation. *Journal of the Americanstatistical Association*, *82*(398), 528–540. https://doi.org/10.2307/2289457

Umbach, N., Naumann, K., Hoppe, D., Brandt, H., Kelava, A., & Schmitz, B. (2017). *Package "nlsem" [Computer* software]. https://cran.r-project.org/web/packages/nlsem/nlsem.pdf

Van Buuren, S. (2012). *Flexible imputation of missing data*. Chapman & Hall/CRC Press.

Van Buuren, S., Boshuizen, H. C., & Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, *18*(6), 681–694. https://doi.org/10.1002/(SICI)1097-0258(19990330)18:6<681::AID-SIM71>3.0.CO;2-R

Van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, *76*(12), 1049–1064. https://doi.org/10.1080/10629360600810434

Van Buuren, S., & Groothuis-Oudshoorn, C. G. M. (2011). MICE: Multivariate imputation by chained equations in

R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Van Ginkel, J. R. (2020). Standardized regression coefficients and newly proposed estimators for R^2 in multiply imputed data. *Psychometrika*, *85*(1), 185–205. https://doi.org/10.1007/s11336-020-09696-4

Van Ginkel, J. R., Van der Ark, L. A., Sijtsma, K., & Vermunt, J. K. (2007). Two-way imputation: A Bayesian method for estimating missing scores in tests and questionnaires, and an accurate approximation. *Computational Statistics & Data Analysis*, *51*(8), 4013–4027. https://doi.org/10.1016/j.csda.2006.12.022

Wechsler, D., Coalson, D. L., & Raiford, S. E. (2008). *WAIS-IV technical and interpretive manual*. Pearson.

Yu, L.-M., Burton, A., & Rivero-Arias, O. (2007). Evaluation of software for multiple imputation of semi-continuous data. *Statistical Methods in Medical Research*, *16*(3), 243–258. https://doi.org/10.1177/0962280206074464

Appendix

In this Appendix, we describe the technical details of the data augmentation (DA) algorithm used to impute according to a moderated factor model.

After filling in starting values for the missing data, the DA algorithm consists of two steps which are iteratively carried out until the imputed values are sufficiently far removed from their starting values: The imputation (I) step and the posterior (P) step. More specifically, suppose that an incomplete dataset $\mathbf{X} = [\mathbf{Y}, \mathbf{m}]$ consists of an observed part $\mathbf{X}_{obs}$ and $\mathbf{X}_{mis}$, and $\boldsymbol{\xi}^{(t)}$ is a set of parameter estimates of the imputation model at iteration t ($t = 1, \ldots, T$), in the I-step, at iteration $(t+1)$ the imputed values $\mathbf{X}_{mis}^{(t+1)}$ are drawn from $P(\mathbf{X}_{mis}|\mathbf{X}_{obs}, \boldsymbol{\xi}^{(t)})$. The P-step is meant for taking into account uncertainty about the parameters of the imputation model, which are unknown. In the P-step the set of model parameters $\boldsymbol{\xi}^{(t+1)}$ are drawn from a posterior distribution of $P(\boldsymbol{\xi}|\mathbf{X}_{obs}, \mathbf{X}_{mis}^{(t+1)})$.

In our specific situation, both $\mathbf{m}_{mis}$ and $\mathbf{Y}_{mis}$ need to be imputed in the I-step. At iteration $t+1$, the I-step for $\mathbf{m}_{mis}$ involves sampling $m_{mis,i}^{(t+1)}$ from its model implied distribution conditional on the data vector $\mathbf{y}_i^{(t)}$ at iteration t. To this end, we assume that m_i is standard normally distributed. As a result, the distribution of m_i conditional on the observed data vector $\mathbf{y}_i$ is given by

$$g(m_i|\mathbf{y}_i) = \frac{f(\mathbf{y}_i|m_i) \times h(m_i)}{\int_{-\infty}^{\infty} f(\mathbf{y}_i|m_i) \times h(m_i) dy} \tag{12}$$

where $h(.)$ is the (standard normal) distribution of m_i.

At iteration $t+1$ $\mathbf{m}_{mis,i}^{(t+1)}$ is sampled from $g(.)$, to replace the missing values in $\mathbf{m}$ using rejection sampling. Rejection sampling involves sampling a proposal value, m_i^*, from a proposal distribution, and accepting this value if $g(m_i^*|\mathbf{y}_i)/[C \times k(m_i^*)] > u$, where u is a random uniform number between 0 and 1, $k(.)$ is the proposal distribution, and C is a tuning parameter between 1 and ∞ which ensures that $g(.) \leq C \times k(.)$.

As proposal distribution we use a normal distribution with mean $E(m_i|\mathbf{y}_i)$ and variance $VAR(m_i|\mathbf{y}_i)$ which are given by

$$E(m_i|\mathbf{y}_i) = \frac{\int_{-\infty}^{\infty} m_i f(\mathbf{y}_i|m_i) \times h(m_i) dm}{\int_{-\infty}^{\infty} f(\mathbf{y}_i|m_i) \times h(m_i) dm} \tag{13}$$

$$VAR(m_i|\mathbf{y}_i) = \frac{\int_{-\infty}^{\infty} m_i^2 f(\mathbf{y}_i|m_i) \times h(m_i) dm}{\int_{-\infty}^{\infty} f(\mathbf{y}_i|m_i) \times h(m_i) dy} - E(m_i|\mathbf{y}_i)^2. \tag{14}$$

For C we use 2 which turns out to work well in practice. Note that approximating $g(m_i|\mathbf{y}_i)$ by a normal distribution (with $E(m_i|\mathbf{y}_i)$ and $VAR(m_i|\mathbf{y}_i)$) and sampling directly from this distribution is not recommendable as due to the non-linear and heteroscedastic nature of the present model, $g(m_i|\mathbf{y}_i)$ can be bimodal and skewed.

Given how the distribution $g(m_i|\mathbf{y}_i)$ is obtained, we can show the I-step of our algorithm. Let $\boldsymbol{\mu}_y^{(t)}|m_i^{(t)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{y}}^{(t)}|m_i^{(t)}$ denote the model predicted mean vector and the model predicted covariance matrix of $\boldsymbol{y}_i$ conditional on $m_i^{(t)}$ at iteration t, respectively. The I-step is as follows:

1. $m_{mis,i}^{(t+1)} \sim g(m_{mis,i}^{(t+1)}|\mathbf{y}_i^{(t)})$.
2. $\mathrm{y}_{mis,ij}^{(t+1)} \sim N(\mu_{y,j}^{(t)}|m_i^{(t)}, \mathbf{y}_{i,-j}; \boldsymbol{\Sigma}_{\boldsymbol{y},jj}^{(t)}|m_i^{(t)}, \boldsymbol{y}_{i,-j})$

where $\mu_{y,j}^{(t)}|m_i^{(t)}, \mathbf{y}_{i,-j}$ is the model predicted mean on variable j conditional on $m_i^{(t)}$ and on the vector observed variables without variable j. Similarly, $\boldsymbol{\Sigma}_{\mathbf{y},jj}^{(t)}|m_i^{(t)}, \mathbf{y}_{i,-j}$ is the variance of variable j conditional on $m_i^{(t)}$ and on the vector of observed variables without variable j. These quantities follow straightforwardly from conditioning a multivariate normal distribution on all but one of its variables (e.g., $\mathrm{X}_1 \mid \mathrm{X}_2, \mathrm{X}_3$).

Next, we will describe the P-step in our algorithm. Define $\upsilon^{(t)} = \left[(\upsilon_{02}^{(t)}, \upsilon_{12}^{(t)}), \ldots, (\upsilon_{0J}^{(t)}, \upsilon_{1J}^{(t)})\right]$ as a vector of intercept parameters, $\boldsymbol{\lambda}^{(t)} = \left\{\left[(\lambda_{011}^{(t)}, \ldots, \lambda_{0J1}^{(t)}), (\lambda_{111}^{(t)}, \ldots, \lambda_{1J1}^{(t)})\right], \ldots, \left[(\lambda_{011}^{(t)}, \ldots, \lambda_{0JD}^{(t)}), (\lambda_{111}^{(t)}, \ldots, \lambda_{1JD}^{(t)})\right]\right\}$ as the loading parameters, $\boldsymbol{\theta}^{(t)} = \left[(\theta_{01}^{(t)}, \theta_{11}^{(t)}), \ldots, (\theta_{0J}^{(t)}, \theta_{1J}^{(t)})\right]$ as the residual variance parameters, $\boldsymbol{\kappa}^{(t)}$ as the mean vector of the factor scores, $\boldsymbol{\varphi}^{(t)} =$

$\left[(\varphi_{01}^{(t)}, \varphi_{11}^{(t)}), \ldots, (\varphi_{0D}^{(t)}, \varphi_{1D}^{(t)})\right]$ as the variance parameters of the factor scores, and $\boldsymbol{\rho}^{(t)} = \left[(\rho_{12}, \ldots, \rho_{1K}), \ldots, (\rho_{j,(j+1)}, \ldots, \rho_{jK}) \ldots, \rho_{(K-1),K}\right]$ as the covariance parameters of the factors scores at iteration t. The total set of parameters at iteration t is $\boldsymbol{\xi}^{(t)} = \left[\boldsymbol{\nu}^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}, \boldsymbol{\kappa}^{(t)}, \boldsymbol{\varphi}^{(t)}, \boldsymbol{\rho}^{(t)}\right]$, and $\boldsymbol{\Psi}^{(t)}$ is its covariance matrix. Finally, let $\boldsymbol{\Lambda}_0^{(t+1)}$ and $\boldsymbol{\Lambda}_1^{(t+1)}$ be two $J \times D$ matrices containing the loading parameters $\lambda_{0jd}^{(t)}$ and $\lambda_{1jd}^{(t)}$, respectively, at iteration t.

The P-step in FS-MFA is done as follows:

1. $\boldsymbol{\Psi}^{(t+1)} \sim W^{-1}\left(N, N\tilde{n}\boldsymbol{\Psi}^{(t)}\right)$
2. $\boldsymbol{\xi}^{(t+1)} \sim MVN(\boldsymbol{\xi}^{(t)}; \boldsymbol{\Psi}^{(t+1)})$
3. $\boldsymbol{\mu}_i^{(t+1)} | m_i^{(t+1)} = \nu_i(m_i)^{(t+1)} + \sum_{d=1}^{D} \lambda_{jd}(m_i)^{(t+1)} \kappa_d(m_i)^{(t+1)}$
   $\nu_i(m_i)^{(t+1)} = \nu_{0j}^{(t+1)} + \nu_{1j}^{(t+1)} m_i^{(t+1)}$
   $\lambda_{jd}(m_i)^{(t+1)} = \lambda_{0jd}^{(t+1)} + \lambda_{1jd}^{(t+1)} m_i^{(t+1)}$
   $\kappa_d(m_i)^{(t+1)} = \kappa_{0d} + \kappa_{1d} m_i^{(t+1)}$
4. $$\boldsymbol{\Sigma}_{\mathbf{y}_i}^{(t+1)} | m_i^{(t+1)} = \boldsymbol{\Lambda}_0^{(t+1)} \boldsymbol{\Sigma}(m_i)^{(t+1)} \boldsymbol{\Lambda}_1^{(t+1)\prime} + diag\left[\theta_1(m_i)^{(t+1)}, \ldots, \theta_J(m_i)^{(t+1)}\right]$$
$$\boldsymbol{\Sigma}(m_i)^{(t+1)} = \begin{bmatrix} \sigma_{11}^2(m_i)^{(t+1)} & & \\ \vdots & \ddots & \\ \sigma_{D1}(m_i)^{(t+1)} & \cdots & \sigma_{DD}(m_i)^{(t+1)} \end{bmatrix}$$
$$\sigma_{dd}^2(m_i)^{(t+1)} = \exp\left(\varphi_{0d}^{(t+1)} + \varphi_{1d}^{(t+1)} m_i^{(t+1)}\right)$$
$$\sigma_{dk}(m_i)^{(t+1)} = \sqrt{\sigma_{dd}^2(m_i)^{(t+1)} \sigma_{kk}^2(m_i)^{(t+1)}} \frac{\exp\left[\rho_{dk}(m_i)^{(t+1)}\right] - 1}{\exp\left[\rho_{dk}(m_i)^{(t+1)}\right] + 1}, (d \neq k)$$
$$\rho_{dk}(m_i)^{(t+1)} = \rho_{0dk}^{(t+1)} + \rho_{1dk}^{(t+1)} m_i^{(t)}$$

For the imputation of $\mathbf{Y}_{mis}$, the procedure is similar to MI under the multivariate normal model, as described by Schafer (1997, chap. 5). The difference with our procedure is that in Schafer's procedure in the P-step, 1) $\boldsymbol{\mu}_i^{(t+1)}$ and $\boldsymbol{\Sigma}_{\mathbf{y}_i}^{(t+1)}$ are the same for all respondents as they do not depend on a moderator variable (but in our procedure they dependent on the moderator), and 2) they are directly drawn from their posterior distributions rather than constructed from drawn parameters from a moderated factor model.