Multilevel exploratory factor analysis of discrete data

Article in Netherlands journal of psychology · December 2012

Suzanne Jak
University of Amsterdam

Marieke E. Timmerman
University of Groningen

Multilevel exploratory factor analysis of discrete data

Exploratory factor analysis (EFA) can be used to determine the dimensionality of a set of items. When data come from clustered subjects, such as pupils within schools or children within families, the hierarchical structure of the data should be taken into account. Standard multilevel EFA is only suited for the analysis of continuous data. However, with the robust weighted least squares estimation procedures that are implemented in the computer program M*plus*, it has become possible to easily conduct EFA of multilevel discrete data. In the present paper, we show how multilevel EFA can be used to determine the dimensionality in discrete two-level data. Measurement invariance across clusters implies equal dimensionality across levels. We describe two procedures, one with and one without measurement invariance restrictions across clusters. Data from educational research serve as an illustrative example.

Where: Netherlands Journal of Psychology, Volume 67, 114-121

Received: 30 August 2012; Accepted: 16 January 2013

Keywords: Exploratory factor analysis; Dimensionality; Discrete data; Multilevel data; Weighted least squares estimation

Authors: Mariska (M.T.) Barendse*, Frans J. Oort**, Suzanne Jak** and Marieke E. Timmerman*

*Heymans Institute of Psychology, University of Groningen, the Netherlands, **Research Institute of Child Development and Education, University of Amsterdam, the Netherlands

Correspondence to: F.J. Oort, Research Institute of Child Development and Education, University of Amsterdam; Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, the Netherlands, e-mail: f.j.oort@uva.nl

Introduction

The dimensionality of a set of items can be defined as the minimum number of underlying unobserved (latent) variables that is needed to describe all relationships between all item responses (Lord & Novick, 1968; Zhang & Stout, 1999). If we restrict ourselves to linear relationships, then exploratory factor analysis (EFA) can be used to assess how many latent variables (or common factors) are needed to explain all item responses (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999; Conway & Huffcutt, 2003). EFA is an appropriate technique to determine dimensionality because the EFA model is unconstrained, so that any misfit can only be attributed to the number of factors being too small. However, ordinary EFA is only suited for the analysis of normally distributed continuous item responses.

Item responses are generally discrete. Test items are often scored as ‘right’ or ‘wrong’, with binary codings 1 and 0. Or respondents give judgements on, for example, a three-point response scale with ‘not applicable to me’, ‘somewhat applicable to me’, and ‘applicable to me’ scored as 1, 2, 3. Wirth and Edwards (2007) give an overview of estimation methods that can be used with discrete item responses. Some of these have been implemented in structural equation modelling computer programs such as M*plus* (Muthén & Muthén, 2010), and so it has become feasible to conduct factor analysis of discrete variables. Barendse, Oort, and Timmerman (2012) conducted a simulation study of EFA of discrete variables and found that robust weighted least squares estimation with polychoric correlations worked well in assessing dimensionality.

In social and behavioural research, we often encounter hierarchically structured data, such as data from students in schools, children in families, or patients sharing the same physicians. Mixed model or multilevel analysis accounts for the dependencies in multilevel data (Snijders & Bosker, 1999). In the case of two-level data, the first level pertains to within-cluster variation (e.g., differences between students within schools) and the second level to between-cluster variation (e.g., differences between schools). Due to the work of Asparouhov and Muthén (2007), the robust weighted least squares estimation implemented in M*plus* (Muthén & Muthén, 2010) can handle multilevel discrete data.

The purpose of this paper is to show how multilevel EFA analysis can be used to assess the dimensionality of a set of discrete responses. We will describe two procedures. In the first procedure we separately assess the dimensionality of within-cluster variance and between-cluster variance, without any restrictions across levels. In the second procedure we assume measurement invariance across clusters, to make sure that the common factors have the same interpretation across clusters. Jak, Oort, and Dolan (2012a) have shown that this measurement invariance restriction implies measurement invariance across levels as well.

Both procedures will be illustrated with data from educational research on student-teacher relationships.

Methods

Below we briefly describe the two-level EFA model, the identification and estimation of its parameters, the evaluation of fit, the two procedures to assess dimensionality, and the rotation of a two-level EFA solution. We currently apply two-level EFA to discrete item responses, but the approach can also be applied to other variables (e.g., continuous scores, counts), and be extended to more than two levels.

With discrete data we assume that the observed discrete item responses are representations of continuous unobserved responses. That is, the vector of observed discrete item responses x_{ij} of individual i in cluster j is considered to be a representation of a vector of underlying continuous response variables y_{ij}, with associated thresholds that determine the x_{ij} values (e.g., Olsson, 1979; Muthén, 1984).

Model

In multilevel models, the underlying continuous variables y_{ij} are decomposed into cluster means μ_j, and individual deviations from the cluster means η_{ij}:

$$y_{ij} = \mu_j + \eta_{ij} \,. \tag{1}$$

The individual deviations η_{ij} are assumed to be independent of the cluster means μ_j so that variance-covariance matrix of y, denoted Σ_{TOTAL} (with variances and covariances across all clusters), is the sum of the variance-covariance matrix of μ, denoted Σ_{BETWEEN} (with variances and covariances between clusters), and the variance-covariance matrix of η, denoted Σ_{WITHIN} (with variances and covariances within clusters),

$$\Sigma_{\text{TOTAL}} = \Sigma_{\text{BETWEEN}} + \Sigma_{\text{WITHIN}} \,. \tag{2}$$

In two-level factor analysis, the between and within variance-covariance matrices can be separately modelled as

$$\Sigma_{\text{BETWEEN}} = \Lambda_B \, \Phi_B \, \Lambda_B' + \Theta_B \,, \tag{3}$$
$$\Sigma_{\text{WITHIN}} = \Lambda_W \, \Phi_W \, \Lambda_W' + \Theta_W \,. \tag{4}$$

In Equation 3, Φ_B is the variance-covariance matrix of the common between factors of the cluster means μ, Λ_B is the matrix of factor loadings of the cluster means on these common between factors, and Θ_B is the (diagonal) matrix with residual variances of the cluster means. In Equation 4, Φ_W is the pooled-within variance-covariance matrix of the common within factors of the individual deviations from the cluster means, Λ_W is the pooled-within matrix of factor loadings of the individual deviations on these common within factors, and Θ_W is the (diagonal) pooled-within matrix with residual variances of the individual deviations.

Measurement invariance

If we want to make sure that the interpretation of the common within factors is the same in all clusters, then we have to assume measurement invariance across clusters (i.e., in factor analysis of mean and covariance structures, intercepts and factor loadings of y-variables are the same across clusters; Muthén, 1994; Rabe-Hesketh, Skrondal, & Pickles, 2004; Jak, Oort, & Dolan, 2012a, 2012b). Jak et al. (2012a) explain that measurement invariance across clusters implies equal factor loadings across levels ($\Lambda_W = \Lambda_B = \Lambda$), yielding the following two-level model:

$$\Sigma_{\text{BETWEEN}} = \Lambda \, \Phi_B \, \Lambda' \,, \tag{5}$$
$$\Sigma_{\text{WITHIN}} = \Lambda \, \Phi_W \, \Lambda' + \Theta_W \,, \tag{6}$$

where Λ is a matrix of factor loadings that is equal across all clusters and across the within and between levels, implying that common factors do have the same interpretation across all clusters and across levels. In addition, there is no residual variance at the between level ($\Theta_B = 0$), implying that no other factors than the common factors are affecting the between-level responses (no 'cluster bias', Jak et al., 2012a).

Identification

In ordinary EFA, the (single level) model is identified with sufficient and necessary scaling and rotation constraints such as an identity matrix for the variance-covariance matrix of common factors ($\Phi = I$) and echelon form for the matrix of factor loadings (Λ elements $\lambda_{pk} = 0$ if $p < k$). In two-level EFA (Equations 3 and 4), sufficient constraints are $\Phi_W = I$, $\Phi_B = I$, and echelon form for both Λ_W and Λ_B. However, if we assume measurement invariance ($\Lambda_W = \Lambda_B = \Lambda$, and $\Theta_B = 0$), then we can estimate the variances of the common factors at the between level (i.e., diagonal(Φ_B) free instead of $\Phi_B = I$). In addition, we can choose either

- to estimate the full factor loading matrix instead of having an echelon form (Λ full free instead of Λ echelon), or

- to estimate correlations between the common factors at the within level (diagonal(Φ_W) = I instead of Φ_W = I), or
- to estimate covariances between the factors at the between level (Φ_B symmetrical free instead of Φ_B diagonal free).

Estimation

The computer program M*plus* provides various estimation methods for SEM with discrete data (Muthén & Muthén, 2010), such as the so-called weighted least squares estimation method with a robust mean-and-variance corrected chi-square fit criterion (WLSMV; Muthén, du Toit, & Spisic, 1997), which has been advocated in previous simulation studies (e.g., Beauducel & Herzberg, 2006; Barendse, et al., 2012). Asparouhov and Muthén (2007) developed a method for multilevel data that can be applied to discrete data, using polychoric correlations. To compare nested multilevel models, one should use the estimation method with a mean-corrected chi-square fit criterion (denoted WLSM; rather than the mean-and-variance corrected WLSMV), as only WLSM provides a valid chi-square statistic to test the difference in fit of nested multilevel models (Muthén, 1998-2004; Satorra & Bentler, 2001).

Evaluation of fit

As the evaluation of fit of multilevel models for discrete data is still subject to study, we resort to fit criteria that are commonly applied in structural equation modelling. A significant chi-square test of overall goodness-of-fit indicates that the model does not fit the data (i.e., the hypothesis of exact population fit is rejected). In addition to the chi-square test of exact fit, we can use the root mean square error of approximation (RMSEA) as an index of approximate fit. RMSEA values below 0.08 and 0.05 indicate satisfactory and close fit, respectively (Browne & Cudeck, 1992). We will also report the standardised root mean square residual (SRMSR) and its weighted counterpart (WRMSR), which indicate the difference between the polychoric correlations and the correlations implied by the EFA model. SRMSR values below 0.05 (e.g., Sivo, Fan, Witta, & Willse, 2006) and WRMSR values below 1.0 (Yu & Muthén, 2002) are considered acceptable.

The difference in fit of two hierarchically related models (or nested models) can be tested with the chi-square difference test. We should note, however, that with WLSM estimation, this chi-square difference is subject to a scaling correction and cannot be calculated by simply taking the difference of the two chi-square values that are associated with the fit of two models (Muthén, 1998-2004; Satorra & Bentler, 2001).

Dimensionality assessment

We describe two procedures to determine the dimensionality of two-level data.

Procedure 1. The first procedure has two steps. In the first step, we leave Σ_{BETWEEN} free to be estimated, impose an exploratory factor model on Σ_{WITHIN} (Equation 4), and fit a series of models with increasing numbers of common within factors to determine the minimum number of common within factors that provides good fit. In the second step, we retain the minimum number of common within factors (determined in the first step), and fit a series of models with increasing numbers of common between factors to determine the minimum number of common between factors that provides good fit.

Procedure 1 may yield a different number of between factors than the number of within factors. So, the dimensionality of the between structure may be different from the dimensionality of the within structure. Still, even if the dimensionality is the same across levels, the interpretation of the between factors is different from the interpretation of the within factors as Λ_W and Λ_B are different. Moreover, the interpretation of the factors across clusters is not the same either, as the values of the Λ_W elements are pooled within values. Matrix Λ_W can be interpreted as the average of as many cluster specific Λ matrices as there are clusters. So, in theory, the Λ_W interpretation may not apply to any of the individual clusters at all.

Procedure 2. In Procedure 2 we require measurement invariance across clusters, which implies $\Lambda_W = \Lambda_B$ and $\Theta_B = 0$ (Jak et al., 2012a). With these restrictions, we fit a series of two-level EFA models to Σ_{WITHIN} and Σ_{BETWEEN} as given by Equations 5 and 6, with increasing numbers of common factors, to determine the minimum number of common factors that provides good fit. Due to the measurement invariance restriction, the common factors have the same number and the same interpretations across all clusters and across both levels.

Rotation

Just as in ordinary (single level) EFA, the solution can be rotated to facilitate interpretation. If the solution is obtained through Procedure 1, using the two-level EFA given by Equations 3 and 4, with both Φ_W and Φ_B equal to identity and both Λ_W and Λ_B having echelon form, then the within and between solutions can be rotated separately, in the same way as in ordinary EFA (Browne, 2001; Oort, 2011).
If the solution is obtained through Procedure 2, using the two-level EFA given by Equations 5 and 6, with Φ_W identity, Φ_B free, and Λ echelon, then we preserve the identical interpretation of within and between factors by rotating the within and between structures together. Application of a rotation criterion

as desired to the echelon Λ yields a transformation matrix T, and rotated factor loadings Λ^* and variance-covariance matrices Φ_W^* and Φ_B^*,

$$\Lambda^* = \Lambda\, T\,, \tag{7}$$

$$\Phi_W^* = (T^{-1})(T^{-1})' = (T'\,T)^{-1}\,, \tag{8}$$

$$\Phi_B^* = (T^{-1})\,\Phi_B\,(T^{-1})'\,. \tag{9}$$

See Browne (2001) for a comprehensive explanation of rotation in EFA.

Illustration

As an illustrative example, we apply multilevel EFA to data that were gathered with the student-teacher relationship scale (STRS; Spilt, Koomen & Jak, 2012). We have complete data from 649 teachers who reported about their relationships with two or three children each, 1493 children in total, aged 3 to 12. The 28 items of the STRS are hypothesised to capture three aspects of the student-teacher relationship: closeness, conflict, and dependency. The items have five-point response scales, ranging from 1 (‘definitely does not apply’) to 5 (‘definitely does apply’).

Preliminary analysis

First we check whether the between-level variances and covariances are sufficiently large to warrant a multilevel analysis. Intra-class coefficients of the item responses vary between 0.15 and 0.49. Furthermore, we fitted a Null Model ($\Sigma_{\text{BETWEEN}} = 0$, Σ_{WITHIN} free) to test whether there is between-level variance, and an Independence Model (Σ_{BETWEEN} diagonal, Σ_{WITHIN} free) to test whether there is between-level covariance. Neither model fits the data: Null Model chi-square = 4547.4, df = 389, $p < 0.001$, RMSEA = 0.085; Independence Model chi-square = 4195.0, df = 378, $p < 0.001$, RMSEA = 0.082. As the intra-class coefficients are high and the Null Model and Independence Model do not fit the data, we conclude that these data require a model that accounts for the two-level hierarchical structure of the data.

Procedure 1 within-level results

Table 1 gives the fit results (chi-square, RMSEA,

Table 1 Series of multilevel exploratory factor analyses to determine the dimensionality

Number of within factors	Number of between factors	DF	Chi-square	RMSEA	SRMSR Within	SRMSR Between	WRMSR	Chi-square difference test Chi-square	Chi-square difference test DF	Chi-square difference test Prob.
Series 1										
1	n/a	350	13737.605	0.160	0.169	-	3.122	-	-	-
2	n/a	323	2302.291	0.064	0.059	-	1.061	16647.295	27	0.000
3	n/a	297	827.289	0.035	0.033	-	0.591	619.279	26	0.000
4	n/a	272	576.248	0.027	0.028	-	0.480	181.013	25	0.000
5	n/a	248	464.731	0.024	0.024	-	0.421	88.807	24	0.000
Series 2										
3	1	647	3086.883	0.050	0.033	0.180	1.182	-	-	-
3	2	620	1332.493	0.028	0.033	0.077	0.685	338.595	27	0.000
3	3	594	1300.727	0.028	0.033	0.067	0.665	43.458	26	0.017
3	4	569	1258.966	0.028	0.033	0.061	0.650	44.247	25	0.010
3	5	545	1223.605	0.029	0.033	0.058	0.639	37.789	24	0.036
Series 3										
1	1	755	14252.961	0.109	0.170	0.589	3.483	-	-	-
2	2	726	4018.500	0.055	0.062	0.304	1.677	2004.806	29	0.000
3	3	697	2705.744	0.044	0.048	0.291	1.236	325.442	29	0.000
4	4	668	1923.310	0.035	0.034	0.278	0.920	198.966	29	0.000
5	5	639	1590.571	0.032	0.032	0.244	0.760	125.307	29	0.000

Note: 1493 pupils are rated by 649 teachers.

SRMSR, WRMSR) for three series of two-level EFA models. In the first series, Σ_{BETWEEN} is unrestricted and Σ_{WITHIN} conforms a one-, two-, three-, four-, or five-factor model (as in Equation 4). The chi-square test is consistently significant, indicating that none of the models fits the data exactly. However, the RMSEA indicates satisfactory fit of the two-factor model and close fit of the three-factor model. The SRMSR and WRMSR indices also suggest acceptable fit of the three-factor model. We therefore continue Procedure 1 with three factors at the within level.

Procedure 1 between-level results

In the second series, Σ_{WITHIN} is restricted to a three-factor model (Equation 4), and Σ_{BETWEEN} is restricted to either a one-, two-, three-, four-, or five-factor model (Equation 3). For each of these models, the chi-square test of exact fit is significant, but due to the gain in degrees of freedom, the relative fit is much better than in the first series of models. According to the RMSEA we would select the EFA model with three within factors and two between factors.

The chi-square difference test indicates that exact fit keeps improving with each additional between factor, but only if we test a 5% level of significance. When testing at a 1% level of significance, we would also select the EFA model with three within factors and two between factors, because at 1%, an additional between factor does not significantly improve exact fit. The same model is also suggested by the WRMSR, but the between-level SRMSR does not fall below 0.05 for any of the models.

Procedure 2 measurement invariance results

In the third series of models we impose measurement invariance restrictions and fit two-level EFA models as given by Equations 5 and 6, with increasing numbers of factors. All chi-square tests are significant, thereby rejecting exact fit. The three-factor model is the first model that meets the RMSEA criterion of close fit (RMSEA < 0.05). The same model also meets the SRMSR criterion (SRMSR < 0.05), but only for the within part. The WRMSR criterion (WRMSR < 1.0) suggests a four-factor model, but Φ_B estimates for this model have unreasonably high standard errors.

Relying on the RMSEA index of fit and on the substantive argument that the STRS is supposed to cover three aspects of student-teacher relationships, we prefer the three-factor model.

The three-factor EFA model with measurement invariance restrictions is nested under the three-within three-between factor model without measurement invariance restrictions in the second series. According to the Satorra and Bentler (2001) chi-square difference test, the hypothesis of measurement invariance should be rejected (chi-square difference = 582.7, df = 103, $p < 0.001$). However, as the RMSEA nevertheless indicates close fit for the restricted model as well, we still prefer the measurement invariant EFA model.

Rotation results

A substantive interpretation of the common factors that is valid across all clusters requires measurement invariance. To facilitate the interpretation of the three-factor two-level EFA model with measurement invariance (Equations 5 and 6), we use the oblimin criterion to rotate the solution (Browne, 2001). As student-teacher relationship factors are likely to be correlated, we opted for oblique rotation, rather than orthogonal. Rotation results are given in Table 2.

From Table 2 it appears that almost all conflict, dependency, and closeness items have their highest loadings on the first, second, and third factor. We have therefore named these factors 'Conflict', 'Dependency', and 'Closeness'. Oblique rotation yields correlated factors. The correlations between the factors Conflict and Dependency (0.39 within level and 0.76 between level), and between Conflict and Closeness (-0.40 within level and -0.64 between level) are substantial. Conspicuously, the within-level correlation between Dependency and Closeness is positive (0.17), albeit small, whereas the between-level correlation is negative (-0.23), showing a difference in the sign of the correlations between judgements of pupils on the one hand and judgements by teachers on the other hand. We note that Koomen et al. (2011) found a zero correlation between Dependency and Closeness, but they neglected the two-level structure of the data and conducted a confirmatory factor analysis with simple structure.

Discussion

In this paper, we have proposed and illustrated two EFA procedures to determine the dimensionality of multilevel discrete data. The first procedure does not involve any across level restrictions, leaving room for different within-level and between-level factor solutions. In that case, the within-level factor loadings (Λ_W) should be interpreted as a summary of all possible individual cluster factor loadings. In the second procedure we assume measurement invariance, to make sure that factors have the same interpretation across all clusters. This assumption entails across-level invariance of within-level and between-level factor loadings ($\Lambda_W = \Lambda_B$).

Without the measurement invariance restriction, common factors may not have the same interpretation across clusters, or across levels, giving room to so-called ‘cluster bias’ (Jak et al., 2012a, 2012b). In the presence of cluster bias, differences between test scores are not completely attributable to differences in the trait(s) one intended to measure. In our student-teacher relationships example, different STRS item scores should be fully explained by differences in scores on the common factors that

Table 2 Exploratory factor analysis of 28 items of the Student-Teacher Relationship Scale (STRS; 649 teachers and 1493 pupils): Oblimin rotation of a three-factor two-level model with measurement invariance

Within- and between-factor loadings ($\Lambda_W = \Lambda_B$)

Items	Conflict	Dependency	Closeness
Closeness items			
I share an affectionate, warm relationship with this child	-0.485	0.050	**1.415**
If upset, this child will seek comfort from me	0.080	0.052	**1.131**
This child is uncomfortable with physical affection or touch from me	0.013	-0.038	**0.601**
This child values his/her relationship with me	-0.206	-0.045	**1.200**
When I praise this child, he/she beams with pride	0.190	-0.066	**0.780**
This child is overly dependent on me	-0.442	0.379	**0.706**
This child tries to please me	-0.009	0.014	**1.031**
It is easy to be in tune with what this child is feeling	0.113	0.076	**1.315**
This child openly shares his/her feelings and experiences with me	-0.514	0.107	**1.104**
This child allows himself/herself to be encouraged by me	0.014	0.147	**0.746**
This child seems to feel secure with me	-0.466	-0.067	**1.225**
Conflict items			
This child and I always seem to be struggling with each other	**1.369**	-0.071	-0.127
This child easily becomes angry with me	**1.293**	0.059	0.106
This child feels that I treat him/her unfairly	**1.313**	0.024	-0.118
This child sees me as a source of punishment and criticism	**0.943**	0.240	-0.423
This child remains angry or is resistant after being disciplined	**1.477**	0.022	0.134
Dealing with this child drains my energy	**1.957**	-0.014	0.067
When this child is in a bad mood, I know we're in for a long and difficult day	**1.583**	0.176	0.163
This child's feelings toward me can be unpredictable or can change suddenly	**1.572**	0.133	-0.117
Despite my best efforts, I'm uncomfortable with how this child and I get along	**1.187**	0.176	**-0.838**
This child whines or cries when he/she wants something from me.	**0.644**	**0.713**	-0.160
This child is sneaky or manipulative with me	**0.918**	0.099	-0.390
Dependency items			
This child reacts strongly to separation from me	0.050	**0.672**	0.069
This child is overly dependent on me	-0.373	**1.609**	-0.203
This child asks for my help when he/she really does not need help	0.241	**0.619**	0.099
This child expresses hurt or jealousy when I spend time with other children	0.525	**0.650**	-0.022
This child fixes his/her attention on me the whole day long	0.039	**0.898**	0.209
This child needs to be continually confirmed by me	0.156	**0.685**	0.028

	Within-factor correlations (Φ_W)			Between-factor correlations (Φ_B)		
	Conflict	Dependency	Closeness	Conflict	Dependency	Closeness
Conflict	1.000			0.201		
				(1.000)		
Dependency	0.391	1.000		0.314	0.857	
				(0.757)	(1.000)	
Closeness	-0.402	0.173	1.000	-0.197	-0.114	0.466
				(-0.644)	(-0.228)	(1.000)

* *Correlations are given within parentheses; factor loadings > 0.6 are in bold type set; residual variances (Θ_W) not shown.*

we named Conflict, Dependency, and Closeness. If there is cluster bias then apparently other between factors, such as the sex of the teacher or size of the class, also directly affect the STRS item scores. Cluster bias in item responses would then invalidate comparisons of groups that differ in, for example, teacher sex or class size.

In the illustrative analysis of the STRS data, the hypothesis of measurement invariance in the three-factor two-level EFA is rejected by the chi-square difference test (WLSM chi-square difference = 582.7, df = 103, $p < 0.001$). With higher dimensional models, the hypothesis is rejected as well (four-factor WLSM chi-square difference = 562.8, df = 124, $p < 0.001$; five-factor WLSM chi-square difference = 603.4, df = 143). This suggests that measurement invariance does not really hold (in the population). However, considering the fit criteria that indicate close fit, we still prefer the three-factor measurement invariant EFA model, especially because the measurement invariance restriction is substantively important. Without this restriction we cannot validly interpret the within-level EFA results, and therefore we are willing to sacrifice exact fit for interpretability.

The evaluation of fit of multilevel models to discrete data is still subject to study, with inconclusive results, both in the structural equation modelling of discrete data and in the structural equation modelling of multilevel data. Fit measures of multilevel models express the combined (mis)fit at multiple levels. As there are many more observations at the within level than at the between level, the within level has more influence on the overall fit than the between level. Ryu and West (2009) and Boulton (2011) proposed level-specific fit measures for multilevel structural equation modelling (e.g., SRMSR within and SRMSR between). As yet, it is most sensible not to rely on a single fit criterion, and to take the within- and between-level sample sizes into account.

In the present study we combined the challenges of multilevel data and discrete data. Our example analysis shows that it is possible to conduct EFA with multilevel discrete data, that it yields interpretable results, but that the evaluation of fit is partly subjective.

Acknowledgements

This publication is supported by MAGW open competition grant 400-09-084 from the Netherlands Organization for Scientific Research (NWO). The authors thank Helma Koomen for making her data available for secondary analysis.

References

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

Barendse, M. T., Oort, F. J. & Timmerman, M. E. (2012). Using exploratory factor analysis to determine the dimensionality of discrete responses. Manuscript submitted for publication.

Beauducel, A. & Herzberg P. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least square estimation in confirmatory factor analysis. *Structural Equation Modeling, 13,* 186-203.

Boulton, A. (2011). *Fit Index Sensitivity in Multilevel Structural Equation Modeling*. Unpublished doctoral dissertation, University of Kansas-Kansas.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21,* 230-258.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36,* 111-150.

Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods 6,* 147-168.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272-299.

Jak, S., Oort, F. J., & Dolan, C. V. (2012a). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling,* in press.

Jak, S., Oort, F. J., & Dolan, C. V. (2012b). Measurement bias in multilevel data. *Structural Equation Modeling,* in press.

Koomen, H. M. Y., Verschueren, K., van Schooten, E., Jak, S., & Pianta R. (2011) 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.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115-132.

Muthén, B. (1994), Multilevel covariance structure analysis. *Sociological Methods & Research 22,* 376-398.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript technical report. Retrieved

April 2, 2011, from http://pages.gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf

Muthén, B. (1998-2004). *M-plus Technical Appendices*. Los Angeles, CA: Muthén & Muthén. Retrieved April 5, 2011, http://statmodel.com/download/techappen.pdf

Muthén, B. O., & Muthén, L. K. (2010). *M-plus user's guide: statistical analysis with latent variables*. Los Angeles, CA: Muthén & Muthén.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44,* 443-460.

Oort, F. J. (2011). Likelihood-based confidence intervals in exploratory factor analysis. *Structural Equation Modeling, 18*, 383-396.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika 69,* 167-190.

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

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507-514.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for 'optimal' cut-off properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education, 74,* 267-288.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Spilt, J., Koomen, H.M.Y & Jak, S. (2012). Are boys better off with male and girls with female teachers? A multilevel investigation of measurement invariance and gender match in teacher-student relationship quality. *Journal of School Psychology, 50*, 363 - 378.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12,* 58-79.

Yu, C. Y., & Muthén, B. (2002). *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes* (Technical Report) Los Angeles: University of California at Los Angeles, Graduate School of Education & Information Studies.

Zhang, J., & Stout, W. F. (1999). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika, 64*, 129-152.

MARISKA (M.T.) BARENDSE
PhD candidate at the Department of Psychometrics and Statistics; Heymans Institute of Psychology, University of Groningen. Research interests include: structural equation modelling, exploratory factor analysis, dimensionality of discrete data, and measurement bias.

FRANS J. OORT
Full professor of Methods and Statistics and Director of the Research Institute of Child Development and Education, University of Amsterdam. His research interests include measurement and measurement bias, and statistical modelling, especially non-standard applications of structural equation modelling.

SUZANNE JAK
PhD candidate in Methods and Statistics, Research Institute of Child Development and Education, University of Amsterdam. Her research interests include structural equation modelling, measurement bias, multilevel structural equation modelling and meta-analytical structural equation modelling.

MARIEKE E. TIMMERMAN
Associate professor in the Psychometrics and Statistics group of the Heymans Institute for Psychological Research, University of Groningen. Her research interest is multivariate modelling of data from complex designs, with a special interest in latent variable models and data reduction methods.