



UvA-DARE (Digital Academic Repository)

Using exploratory factor analysis to determine the dimensionality of discrete responses

Barendse, M.T.; Oort, F.J.; Timmerman, M.E

DOI

[10.1080/10705511.2014.934850](https://doi.org/10.1080/10705511.2014.934850)

Publication date

2015

Document Version

Final published version

Published in

Structural Equation Modeling

[Link to publication](#)

Citation for published version (APA):

Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using exploratory factor analysis to determine the dimensionality of discrete responses. *Structural Equation Modeling*, 22(1), 87-101. <https://doi.org/10.1080/10705511.2014.934850>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Using Exploratory Factor Analysis to Determine the Dimensionality of Discrete Responses

M. T. Barendse,¹ F. J. Oort,² and M. E. Timmerman¹

¹University of Groningen, The Netherlands

²University of Amsterdam, The Netherlands

Exploratory factor analysis (EFA) is commonly used to determine the dimensionality of continuous data. In a simulation study we investigate its usefulness with discrete data. We vary response scales (continuous, dichotomous, polytomous), factor loadings (medium, high), sample size (small, large), and factor structure (simple, complex). For each condition, we generate 1,000 data sets and apply EFA with 5 estimation methods (maximum likelihood [ML] of covariances, ML of polychoric correlations, robust ML, weighted least squares [WLS], and robust WLS) and 3 fit criteria (chi-square test, root mean square error of approximation, and root mean square residual). The various EFA procedures recover more factors when sample size is large, factor loadings are high, factor structure is simple, and response scales have more options. Robust WLS of polychoric correlations is the preferred method, as it is theoretically justified and shows fewer convergence problems than the other estimation methods.

Keywords: discrete data, exploratory factor analysis, robust maximum likelihood estimation, weighted least squares estimation

The dimensionality of test data is defined as the minimum number of latent variables that is needed to describe all statistical dependencies in the data (Lord & Novick, 1968; Zhang & Stout, 1999). A correct indication of dimensionality helps to get insight into the structure that underlies the responses to test items. Determination of the dimensionality and the associated structure is essential in the development and the evaluation of tests in behavioral and social sciences. Among methods to assess dimensionality, factor analytic methods seem to be the most popular (e.g., Conway & Huffcutt, 2003; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Preacher & MacCallum, 2003; Ten Holt, van Duijn, & Boomsma, 2010). In factor analysis, linear relations between observed variables are explained by unobserved, latent variables or common factors.

Exploratory factor analysis (EFA) does not constrain the factor structure in any way, and by applying maximum likelihood (ML) estimation, the chi-square measure of overall

goodness-of-fit can be used as a test of dimensionality. However, simulation studies have shown that the chi-square test does not always accurately retrieve the correct number of factors (Beauducel, 2001; Hayashi, Bentler, & Yuan, 2007). Possible explanations are small sample size, nonnormality, zero error variances, and rank deficiency (Hayashi et al., 2007). Dimensionality can also be determined by using the chi-square difference test to compare the fit of two nested models with different numbers of factors, to test whether additional factors improve the fit significantly. Still, the difference test is subject to the same problems as the overall chi-square measures on which it is based, generally resulting in too many factors (Hayashi et al., 2007).

Overfactoring might also be the result of the existence of common variance that is due to factors that are trivial to the substantive test content. Researchers who prefer to disregard “minor factors” might consider the root mean square error of approximation (RMSEA; Browne & Cudeck, 1992) as the fit index of choice to determine dimensionality. The RMSEA can be written as a function of the chi-square measure, as most alternative fit indices. A notable exception is the standardized root mean square residual (SRMSR) that summarizes the differences between fitted and observed

Correspondence should be addressed to F. J. Oort, Department of Education, University of Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, The Netherlands. E-mail: f.j.oort@uva.nl

correlations. As far as we know, the SRMSR has never been considered in studies of dimensionality assessment.

Factor analysis by maximizing the likelihood of covariances (or correlations) is really only justified when analyzing continuous responses. However, most tests in the behavioral and social sciences consist of items with discrete response scales. ML analysis of covariances or correlations between discrete responses is often applied but yields biased parameters as well as incorrect standard errors and fit statistics (Dolan, 1994; Johnson & Creech, 1983; Muthén & Kaplan, 1985, 1992; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Robust maximum likelihood (MLR) methods take violations of multivariate normality into account by adjusting standard errors and fit indices (Yuan & Bentler, 2000; Muthén & Muthén, 2010), but are not suited for discrete data either. ML analysis of polychoric correlations reproduces the measurement model correctly, and yields accurate and consistent parameter estimates, but incorrect standard errors and test statistics (Dolan, 1994; Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2010).

Wirth and Edwards (2007) reviewed methods that take the discrete nature of test items into account, one of which is weighted least squares (WLS; Browne, 1982, 1984) factor analysis of polychoric correlations. WLS analysis with the full weight matrix of asymptotic variances and covariances requires very large sample sizes to obtain accurate results (Dolan, 1994; Muthén & Kaplan, 1992; Rigdon & Ferguson, 1991). Therefore, modified WLS methods have been suggested, estimating the model parameters by using the diagonal of the weight matrix only, and subsequently adjusting the chi-square measure and standard errors (Satorra & Bentler, 1994; Yuan & Bentler, 1998). One such method is the robust WLS method suggested by Muthén, du Toit, and Spisic (1997), which is implemented in the *Mplus* computer program (Asparouhov & Muthén, 2010; Muthén & Muthén, 2010) and has been shown to give accurate results in a simulation study of confirmatory factor analysis (Beauducel & Herzberg, 2006; Flora & Curran, 2004; Yang-Wallentin, Jöreskog, & Luo, 2010).

The purpose of this article is to investigate the usefulness of factor analysis to assess the dimensionality of discrete test responses. We generate data under various conditions, considering both major and minor factors, and apply EFA with various estimation methods and fit criteria.

METHOD

We apply EFA to simulated continuous and discrete data. We generate item responses with various response scales (continuous, dichotomous, three-point, and four-point), and we vary the size of the factor loadings (high, low), the factor structure (simple, complex), and sample size (small, large). In each condition, 1,000 data sets are generated and analyzed with five different estimation methods (based on

ML and WLS), using different criteria to determine the number of common factors (the chi-square test, the chi-square difference test, the RMSEA, and the RMSR). The performance of estimation methods and fit criteria in the different conditions is evaluated by comparing model selection rates.

Data Generation: Continuous Responses

Continuous responses. Data are generated with a model that is representative for models regularly used in empirical studies and, similar to the model used in the simulation study of Olsson, Troye, and Howell (1999), we choose a common factor model for 12 observed variables, with 3 major common factors and 4 minor common factors, as depicted in Figure 1.

Continuous responses to 12 items are generated according to the linear model,

$$x = \tau + \Lambda \xi + \varepsilon, \quad (1)$$

where, for an arbitrary subject, x is a 12×1 vector of item responses, ξ is a 7×1 vector of common factor scores, ε is a 12×1 vector of residual factor scores, τ is a 12×1 vector of intercepts, and Λ is a 12×7 matrix of common factor loadings. In all conditions, intercept values are chosen

$$\tau = \begin{bmatrix} 0.0 \\ 0.2 \\ 0.5 \\ 0.8 \\ 0.0 \\ 0.2 \\ 0.5 \\ 0.8 \\ 0.0 \\ 0.2 \\ 0.5 \\ 0.8 \end{bmatrix}. \quad (2)$$

In conditions with high factor loadings on the major factors, factor loadings are chosen:

$$\Lambda = \begin{bmatrix} \sqrt{0.5} & 0.0 & 0.0 & 0.2 & 0.0 & 0.0 & 0.0 \\ \sqrt{0.5} & 0.0 & 0.0 & 0.0 & 0.2 & 0.0 & 0.0 \\ \sqrt{0.5} & 0.0 & 0.0 & 0.0 & 0.0 & 0.2 & 0.0 \\ \sqrt{0.5} & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.2 \\ 0.0 & \sqrt{0.5} & 0.0 & 0.2 & 0.0 & 0.0 & 0.0 \\ 0.0 & \sqrt{0.5} & 0.0 & 0.0 & 0.2 & 0.0 & 0.0 \\ 0.0 & \sqrt{0.5} & 0.0 & 0.0 & 0.0 & 0.2 & 0.0 \\ 0.0 & \sqrt{0.5} & 0.0 & 0.0 & 0.0 & 0.0 & 0.2 \\ 0.0 & 0.0 & \sqrt{0.5} & 0.2 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & \sqrt{0.5} & 0.0 & 0.2 & 0.0 & 0.0 \\ 0.0 & 0.0 & \sqrt{0.5} & 0.0 & 0.0 & 0.2 & 0.0 \\ 0.0 & 0.0 & \sqrt{0.5} & 0.0 & 0.0 & 0.0 & 0.2 \end{bmatrix}, \quad (3)$$

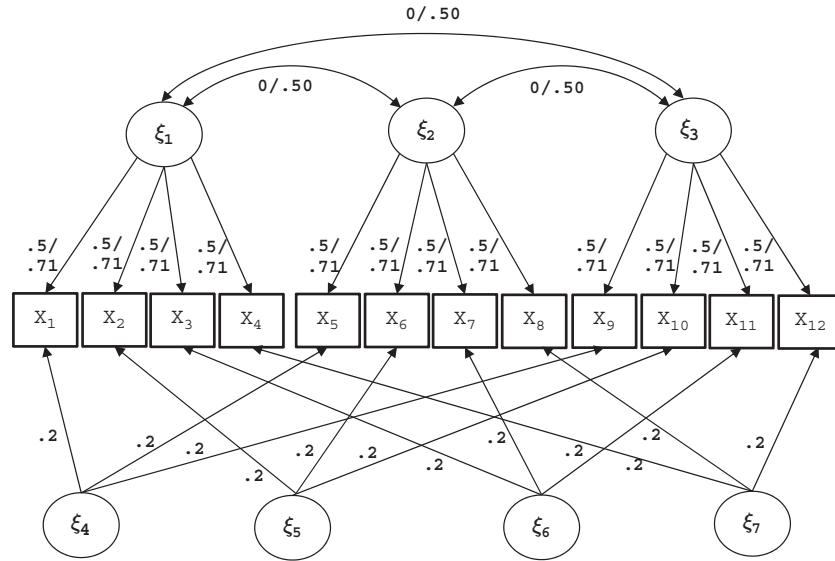


FIGURE 1 Data generation model. *Note.* To improve intelligibility, residual factors are omitted; slashes in 0.0/0.5 and 0.5/ $\sqrt{0.5}$ indicate that parameters take on different values in different conditions.

and in conditions with small factor loadings, the $\sqrt{0.5}$ values are replaced by 0.5 values. Common factor scores are drawn from a multivariate normal distribution with zero means and a 7×7 variance–covariance matrix Φ . In small sample conditions, we draw $200 \times 7 \xi$ values, and in large sample conditions we draw $1,000 \times 7 \xi$ values. In simple structure conditions, Φ is an identity matrix, and in complex structure conditions we choose

$$\Phi = \begin{bmatrix} 1.0 & & & & & & & \\ 0.5 & 1.0 & & & & & & \\ 0.5 & 0.5 & 1.0 & & & & & \\ 0.0 & 0.0 & 0.0 & 1.0 & & & & \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & & & \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & & \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}. \quad (4)$$

Note that the common factors can be rotated to uncorrelated factors, provided that the loadings are counterrotated, losing the simple Λ structure of Equation 3. Values for the residual factor scores, 200×12 in small sample conditions and $1,000 \times 12$ in large sample conditions, are drawn from a multivariate normal distribution with zero means and a 12×12 diagonal variance–covariance matrix Θ with

$$\Theta = I - \text{diag}(\Lambda \Phi \Lambda'). \quad (5)$$

With our choice of parameter values, and with the variance–covariance matrix of the observed variables given by

$$\Sigma = \Lambda \Phi \Lambda' + \Theta, \quad (6)$$

the expected variances of the observed variables are one. For each of the observed variables, the major factors explain

50% of the variance in the conditions with large factor loadings ($\sqrt{0.5}$) and 25% in conditions with low factor loadings (0.5). The minor factors explain only 4% of the variance in all conditions. The misspecification error of a three-factor model that disregards the minor factors can be considered approximation error. The fixed RMSR can be used as an index of this approximation error. In our case it equals 0.0157, which is somewhat smaller than the approximation errors of the models considered by Olsson et al. (1999).

People who are not willing to accept any approximation error should expect a model with six common factors to fit exactly. Six common factors, not seven, because although we generate data with a seven common factor model, the dimensionality of Σ is six, as with our choices of Λ and Φ , the rank of $\Lambda \Phi \Lambda'$ equals six, so six common factors suffice to describe the dependencies between the observed variables.

Discrete responses. Discrete data are generated by categorizing the continuous item responses into discrete item responses. We consider two-, three-, and four-point response scales, as Dolan (1994) showed that ML factor analysis of five-point responses already gives results that are similar to the analysis of continuous responses.

Zero is the cut value that is chosen to dichotomize the continuous item responses, yielding expected proportions of [0.50, 0.50], [0.42, 0.58], [0.31, 0.69], and [0.21, 0.79] for items with intercepts of 0.0, 0.2, 0.5, and 0.8. Three-point responses are generated with cut values of -0.44 and 0.44 , yielding expected proportions of [0.33, 0.34, 0.33], [0.26, 0.33, 0.41], [0.17, 0.30, 0.52], and [0.11, 0.25, 0.64], and four-point responses are generated with cut values of -0.67 , 0 , and 0.67 , yielding expected proportions of [0.25, 0.25, 0.25, 0.25], [0.19, 0.23, 0.26, 0.32], [0.12, 0.19, 0.26, 0.43], and [0.07, 0.14, 0.24, 0.55].

Data Analysis

Combining all variations in a fully crossed design gives 32 conditions. That is, two sizes of factor loadings (high, low) \times two factor structures (simple, complex) \times two sample sizes (200, 1,000) \times four response scales (continuous, dichotomous, three-point, four-point). With 1,000 replications in each condition, we have a total of 32,000 data sets.

Exploratory factor analysis. To each of the 32,000 data sets we fit exploratory factor models with one through seven common factors. Identification is achieved by choosing commonly used scaling and rotation constraints. So, different from the data generation model given by Equation 6, we substitute an identity matrix for Φ and we choose an echelon form for Λ , fixing its upper triangle at zero ($\lambda_{ij} = 0$ for all $i < j$) and leaving all other factor loadings free to be estimated (λ_{ij} free for all $i \geq j$).

We apply five estimation methods: ML of covariances, ML of polychoric correlations, MLR, WLS of polychoric correlations, and robust WLS (WLSMV) of polychoric correlations, as implemented in the computer program *Mplus* (Muthén & Muthén, 2010). We consistently used the *Mplus* ESEM procedure, except for ML of polychoric correlations, where it appeared necessary to additionally constrain the residual variance to be positive.

ML estimation is applied to all data sets in all conditions, although its use is justified only in conditions with continuous, normally distributed responses. The MLR estimation method takes nonnormality into account, yielding robust standard errors and a scaling correction for the chi-square statistic (Yuan & Bentler, 2000), but its use with discrete data is not justified either. However, for the purpose of comparison, we apply the MLR estimation method in conditions with three-point and four-point responses (dichotomous data contain too little information for MLR estimation). We also apply ML to the polychoric correlations between the discrete responses. The WLS and WLSMV estimation methods are best suited for the analysis of discrete data, as they are applied to tetrachoric and polychoric correlations, using the asymptotic variances and covariances as a weight matrix. WLS uses a full weight matrix for the estimation of parameters, standard errors, and chi-square test, whereas WLSMV uses a diagonal weight matrix for parameter estimation and a full weight matrix to obtain standard errors and a mean and variance-adjusted chi-square test statistic (Asparouhov & Muthén, 2010; Muthén & Muthén, 2010).

Fit criteria. To evaluate the fit of each model to each data set we use the chi-square test of overall goodness-of-fit at a 5% level of significance. For comparison, we will also use the RMSEA to evaluate fit. RMSEA values below 0.05 are usually considered indications of close fit (Browne & Cudeck, 1992; MacCallum, Browne, & Sugawara, 1996),

but Hu and Bentler (1999; Yu & Muthén, 2002) suggested .06 as a cutoff, and we also consider some other RMSEA cutoff values.

Most other fit indices are highly correlated with the chi-square statistic and the RMSEA, and are therefore not considered in this study. An exception might be the SRMSR that summarizes the differences between observed and fitted correlations. Hu and Bentler (1999) suggested .08 as an SRMSR cutoff value and Sivo, Fan, Witt, and Willse (2006) suggested .05. In our study, such cutoff values appear too lenient and do not discriminate between the fitted models in different conditions, so instead we choose .04 as the SRMSR cutoff value above which we reject model fit. With the WLS estimation method, the SRMSR is replaced by a weighted root mean square residual (WRMSR; Muthén, 1998–2004). Cutoff values for WRMSR were discussed by Yu and Muthén (2002), who suggested cutoff values as high as .9 and 1.0. However, in our study we find that these values are much too high as cutoffs, and that a WRMSR cutoff value of .5 better discriminates between the fit of different models.

Model selection. After evaluating the fit of all models to all data sets and calculating the rejection rates for the estimation methods and fit criteria described earlier, we follow two selection procedures to determine the dimensionality of each data set. The first procedure is to go through the models, starting with the one-factor model, then the two-factor model, and so on, through the seven-factor model, and select the first model that fits the data, according to the previously mentioned fit criteria.

The second procedure is to go through the models and find the last model that shows significant improvement of fit, according to a significant chi-square difference test at a 5% level of significance. With ML and WLS estimation the chi-square difference test is just the difference between the chi-square values of the two nested models. With MLR estimation, however, the chi-square difference and its scaling correction are a function of the chi-square values and scaling corrections of the nested models (Satorra & Bentler, 2001). With WLSMV estimation, the chi-square difference is subject to a mean and variance correction as described by Asparouhov and Muthén (2006; Asparouhov & Muthén, 2010), similar to the correction of the chi-squares of the individual models.

Evaluation. The performance of the various estimation methods and fit criteria under different conditions is evaluated by determining the numbers of data sets for which the models of increasing dimensionality are selected. Logistic regression analyses are used to compare the results. We expect the chi-square test of (exact) fit to select models with larger numbers of common factors than the RMSEA index that allows for approximation error and that might not find minor factors.

RESULTS

We first give the full ML results for continuous responses. These results can serve as a benchmark when we subsequently summarize the results of estimation methods for discrete responses.

Continuous Responses

Table 1 gives rejection and selection rates for factor models with one through seven dimensions, for each of the eight conditions. The NC column shows the number of data sets with which the ML method converged to a solution. We note that convergence is almost always achieved with one-, two-, and three-factor models, but that convergence becomes very difficult with models with more factors, especially when sample sizes are small (as in Conditions 1–4).

Next, Table 1 gives the rejection rates according to significant chi-square values ($\alpha = .05$), RMSEA values larger than .05, and SRMSR values larger than .04. These rejection rates are calculated both as proportions of successful analyses (NR/NC) and as proportions of the total number of analyses (NR/NT). The two proportions do not differ much, once more indicating that nonconvergence is caused by trying to fit models with redundant factors.

As an aside, we note that the rejection rates according to chi-square and RMSEA values are close to the expected rejection rates that can be calculated on the basis of the noncentral chi-square distribution, but only for the three-factor model, and only after correcting the fit results by $(N - 1)/N$ (as *Mplus* uses N as the multiplier of the discrepancy function). Models with fewer factors do not meet the regularity assumptions for calculating the noncentrality parameter (Steiger, Shapiro, & Browne, 1985). For models with more than three factors, the rejection rates are biased because of frequent convergence problems. It should be noted that the estimation of the noncentrality parameter could be problematic anyway (Olsson, Foss, & Breivik, 2004).

Finally, Table 1 gives the selection rates for each model, according to the procedure of selecting the first model that fits, using the chi-square, RMSEA, and SRMSR fit criteria, and the procedure of selecting the last model that significantly improves fit, using the chi-square difference test ($\alpha = .05$). As expected, the procedures that use the chi-square test generally select models with larger numbers of factors than procedures that use the RMSEA index. In addition, models with larger numbers of factors are more frequently selected in conditions with large sample size, high factor loadings, and simple structure than in conditions with small sample size, low factor loadings, and complex structure.

Of course, rejection and selection rates primarily depend on the arbitrary choices of a level of significance for the chi-square test and cutoffs for the RMSEA and SRMSR indices.

Discrete Responses

Discrete data have been analyzed with five estimation methods: ML of covariances, ML of polychoric correlations, MLR, WLS, and WLSMV. Full results can be downloaded from the website of the corresponding author.¹ Here we summarize the results in two ways. First, we consider rejection and selection rates across all conditions with discrete data (in Tables 2 and 3, to be discussed later). Second, we examine differences in model selection between conditions and estimation methods, using ordinal logistic regression analyses (Tables 4 and 5). When summarizing the results, we only consider the results of ML of covariances, WLS, and WLSMV. We refrain from further presenting the results of MLR and ML of polychoric correlations, because they show problematic behavior in this study. With MLR estimation, the rejection rates are consistently above zero, even for models with large numbers of factors, due to an inconsistent estimate of the MLR scaling correction factor. This inconsistency also causes frequent negative chi-square differences. Furthermore, MLR estimation is associated with the highest proportions of nonconvergence of all estimation methods (Table 6). Moreover, the selection procedure that relies on the chi-square difference test often fails to select any model (e.g., in 34% of the small sample cases and in 48% of the large sample cases) with MLR. With ML of polychoric correlations, the rejection rates are unreasonably high. As a result, selection procedures that rely on the chi-square test fail to select any model in the majority of small sample size cases (see footnote 1).

Rejection rates. Table 2 gives the rejection rates for ML, WLS, and WLSMV analyses with various fit criteria, for small sample conditions and large sample conditions separately, but across all other conditions, and across dichotomous, three-point, and four-point response scales, totaling 12,000 data sets for each estimation method. Numbers and proportions of data sets with which convergence was successful are given in the NC and NC/NT columns.

With ML analysis of covariances, the convergence percentages for the four-factor model are just 41.2% in small sample conditions and 72.2% in large sample conditions, and even worse for models with more factors. It appears that with the common .05 cutoff choice for the RMSEA, Model 3 is rejected in 8.3% of the small sample cases and in only 0.1% of the large sample cases, whereas the chi-square test is significant at the 5% level in 0.1% and 55.0% of the small and large sample cases, respectively. In contrast with the chi-square tests that gain power with larger sample sizes,

¹Tables with full results of all estimation methods and fit criteria can be downloaded from the website of the corresponding author at <http://www.uva.nl/over-de-uva/organisatie/medewerkers/content/o/o/f.j.oort/f.j.oort.html> or <http://tinyurl.com/d7e7c64> (under Miscellaneous: Exploratory Factor Analysis of Discrete Data).

TABLE 1
Model Convergence and Model Rejection and Selection Rates for Maximum Likelihood Analysis of Continuous Responses

	df	Model Rejection Rates							Model Selection Rates				
		Convergence		Chi-Square ($\alpha = .05$)		RMSEA (Cut = .05)		SRMSR (Cut = .04)		Chi-Square ($\alpha = .05$)	RMSEA (Cut = .05)	SRMSR (Cut = .04)	Difference Test ($\alpha = .05$)
		NC	NR/NC	NR/NT	NR/NC	NR/NT	NR/NC	NR/NT	NS/NT	NS/NT	NS/NT	NS/NT	
<i>Condition 1: Simple structure/high factor loadings/sample size = 200</i>													
1-factor model	54	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.000	0.000	0.000	0.000	
2-factor model	43	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.000	0.000	0.000	0.000	
3-factor model	33	1,000	0.369	0.369	0.298	0.298	0.000	0.000	0.631	0.702	1.000	0.352	
4-factor model	24	626	0.053	0.033	0.054	0.034	0.000	0.000	0.199	0.154	0.000	0.405	
5-factor model	16	292	0.000	0.000	0.003	0.001	0.000	0.000	0.026	0.023	0.000	0.078	
6-factor model	9	112	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.003	0.000	0.002	
7-factor model	3	44	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.003	0.000	0.000	
No model selected									0.137	0.115	0.000	0.163	
<i>Condition 2: Complex structure/high factor loadings/sample size = 200</i>													
1-factor model	54	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.000	0.000	0.000	0.000	
2-factor model	43	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.000	0.000	0.000	0.000	
3-factor model	33	1,000	0.344	0.344	0.277	0.277	0.000	0.000	0.656	0.723	1.000	0.381	
4-factor model	24	617	0.058	0.036	0.068	0.042	0.000	0.000	0.174	0.124	0.000	0.384	
5-factor model	16	287	0.007	0.002	0.010	0.003	0.000	0.000	0.030	0.030	0.000	0.072	
6-factor model	9	120	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.010	0.000	0.002	
7-factor model	3	44	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	
No model selected									0.129	0.112	0.000	0.161	
<i>Condition 3: Simple structure/low factor loadings/sample size = 200</i>													
1-factor model	54	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.000	0.000	0.000	0.000	
2-factor model	43	999	0.989	0.988	0.962	0.961	1.000	0.999	0.011	0.038	0.000	0.000	
3-factor model	33	994	0.156	0.155	0.113	0.112	0.076	0.076	0.828	0.844	0.918	0.630	
4-factor model	24	465	0.004	0.002	0.004	0.002	0.000	0.000	0.067	0.044	0.036	0.244	
5-factor model	16	172	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.002	0.001	0.028	
6-factor model	9	89	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.004	0.003	0.000	
7-factor model	3	39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
No model selected									0.086	0.068	0.042	0.098	
<i>Condition 4: complex structure/low factor loadings/sample size = 200</i>													
1-factor model	54	1,000	0.965	0.965	0.882	0.882	1.000	1.000	0.035	0.118	0.000	0.000	
2-factor model	43	947	0.543	0.514	0.378	0.358	0.901	0.853	0.402	0.480	0.094	0.110	
3-factor model	33	813	0.081	0.066	0.057	0.046	0.025	0.020	0.392	0.278	0.720	0.551	
4-factor model	24	388	0.003	0.001	0.005	0.002	0.000	0.000	0.036	0.021	0.028	0.163	
5-factor model	16	172	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.001	0.009	0.008	
6-factor model	9	64	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.002	0.000	
7-factor model	3	28	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.000	
No model selected									0.129	0.100	0.146	0.168	
<i>Condition 5: Simple structure/high factor loadings/sample size = 1,000</i>													
1-factor model	54	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.000	0.000	0.000	0.000	
2-factor model	43	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.000	0.000	0.000	0.000	
3-factor model	33	1,000	0.997	0.997	0.047	0.047	0.000	0.000	0.003	0.953	1.000	0.002	
4-factor model	24	941	0.892	0.839	0.002	0.002	0.000	0.000	0.099	0.044	0.000	0.026	
5-factor model	16	785	0.321	0.252	0.000	0.000	0.000	0.000	0.451	0.001	0.000	0.240	
6-factor model	9	635	0.005	0.003	0.000	0.000	0.000	0.000	0.253	0.001	0.000	0.454	
7-factor model	3	209	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.000	0.000	0.007	
No model selected									0.178	0.001	0.000	0.271	
<i>Condition 6: Complex structure/high factor loadings/sample size = 1,000</i>													
1-factor model	54	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.000	0.000	0.000	0.000	
2-factor model	43	999	1,000	0.999	1,000	0.999	1,000	0.999	0.000	0.000	0.000	0.000	
3-factor model	33	1,000	0.995	0.995	0.060	0.060	0.000	0.000	0.005	0.940	1.000	0.000	
4-factor model	24	946	0.889	0.841	0.002	0.002	0.000	0.000	0.100	0.057	0.000	0.029	
5-factor model	16	819	0.305	0.250	0.000	0.000	0.000	0.000	0.486	0.001	0.000	0.257	
6-factor model	9	661	0.005	0.003	0.000	0.000	0.000	0.000	0.242	0.001	0.000	0.484	
7-factor model	3	225	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.003	
No model selected									0.156	0.001	0.000	0.227	

(Continued)

TABLE 1
(Continued)

	df	Model Rejection Rates							Model Selection Rates			
		Convergence		Chi-Square ($\alpha = .05$)		RMSEA (Cut = .05)		SRMSR (Cut = .04)	Chi-Square ($\alpha = .05$)	RMSEA (Cut = .05)	SRMSR (Cut = .04)	Difference Test ($\alpha = .05$)
		NC	NR/NC	NR/NT	NR/NT	NR/NC	NR/NT	NS/NT	NS/NT	NR/NT	NS/NT	NS/NT
<i>Condition 7: Simple structure/low factor loadings/sample size = 1,000</i>												
1-factor model	54	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.000	0.000	0.000	0.000
2-factor model	43	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.000	0.000	0.000	0.000
3-factor model	33	1,000	0.773	0.773	0.000	0.000	0.000	0.000	0.227	1.000	1.000	0.066
4-factor model	24	827	0.305	0.252	0.000	0.000	0.000	0.000	0.394	0.000	0.000	0.339
5-factor model	16	515	0.027	0.014	0.000	0.000	0.000	0.000	0.172	0.000	0.000	0.295
6-factor model	9	324	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.074
7-factor model	3	172	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.001
No model selected									0.163	0.000	0.000	0.225
<i>Condition 8: Complex structure/low factor loadings/sample size = 1,000</i>												
1-factor model	54	1,000	1,000	1,000	0.999	0.999	0.999	0.999	0.000	0.001	0.001	0.000
2-factor model	43	999	1,000	0.999	0.583	0.582	0.049	0.049	0.000	0.416	0.949	0.000
3-factor model	33	1,000	0.747	0.747	0.000	0.000	0.000	0.000	0.253	0.583	0.050	0.078
4-factor model	24	778	0.261	0.203	0.000	0.000	0.000	0.000	0.378	0.000	0.000	0.366
5-factor model	16	468	0.017	0.008	0.000	0.000	0.000	0.000	0.133	0.000	0.000	0.251
6-factor model	9	282	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.000	0.000	0.047
7-factor model	3	118	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.000	0.000	0.001
No model selected									0.175	0.000	0.000	0.257

Note. RMSEA = root mean square error of approximation; SRMSR = standardized root mean square residual; NC = number of analyses with convergence; NR = number of data sets with which the model is rejected; NT = total number of data sets (1,000); NS = number of data sets with which the model is selected.

both RMSEA and SRMSR results show smaller rejection rates with larger sample sizes. Even with an SRMSR cutoff as small as .03, Model 3 is never rejected in large sample cases.

The rejection rates with WLS estimation generally show the same picture as those with ML estimation, but with WLS estimation the convergence percentages are lower: 33.0% and 66.7% for the four-factor model in small and large sample conditions.

The WLSMV estimation method has better convergence than the ML and WLS estimation methods: 47.1% and 74.9% percent for the four-factor model in small and large sample conditions. In small sample conditions, the three-factor model is rejected in 5.8% of the cases by the chi-square test at $\alpha = .05$ and in 3.8% of the cases by RMSEA cutoff at .05. In large sample conditions, the chi-square test has more power and rejects the three-factor model in 45.8% of the cases. Yet the three-factor model is never rejected by the .05 RMSEA cutoff in large sample conditions. With the WRMSR index, with a .50 cutoff, Model 3 is rejected in 21.9% of the small sample cases and in 65.0% of the large sample cases.

Selection rates. Table 3 gives the model selection rates for ML, WLS, and WLS estimation with various fit criteria. In the small sample size conditions, both the chi-square test and the RMSEA index select Model 3 most of the time. In the large sample conditions, the RMSEA index selects Model 3 even more frequently than in the small sample conditions,

whereas the chi-square test more frequently points to models with larger numbers of factors. As expected, the chi-square difference test generally selects models with more factors than the stand-alone chi-square test, but it also more often ends up without selecting a model.

Convergence problems hinder model selection, especially in small sample conditions, and especially with the WLS estimation procedure, as it fails to select a model in 30.9% of the cases with the chi-square test ($\alpha = .05$), in 35.5% of the cases with the chi-square difference test ($\alpha = .05$), and in 26.8% of the cases with the RMSEA (cutoff = .05).

The WLS estimation procedure shows the lowest percentages of selection failures. With WLS, the chi-square test at $\alpha = .05$, the .05 RMSEA cutoff, the .50 WRMSR cutoff, and the chi-square difference test at $\alpha = .05$, all select Model 3 in more than 60% of the small sample cases. In the large sample cases, the .50 WRMSR cutoff and the chi-square tests more frequently select models with higher numbers of factors.

Condition effects on model selection. To summarize the effects of different conditions and response scales on model selection, Table 4 gives the results of ordinal logistic regression analyses for each estimation method (ML of covariances, WLS, WLSMV) and five fit criteria (chi-square test with $\alpha = .05$, RMSEA index with cutoff .05, SRMSR index with cutoff .04, WRMSR index with cutoff .50, chi-square difference test with $\alpha = .05$), with the number of factors as the dependent variable, and the manipulated factors as predictors (using dummy coding).

The first part of Table 4 shows the effects on model selection through ML estimation. It appears that with the chi-square test, the odds to select a model of higher dimensionality in conditions with high factor loadings are 12.3 times higher than in conditions with small factor loadings. Likewise, the odds to select a model with an additional factor in conditions with simple structure are 1.8 times higher than in conditions with complex structure, and the odds in large sample size conditions are 42.3 times higher than in small sample size conditions. The odds to select a model with an additional factor in conditions with three-point, four-point, and continuous responses are 3.2, 4.9, and 16.0 times higher than in conditions with dichotomous responses.

The other fit criteria show a similar picture, with the exception of the sample size effect on model selection with the SRMSR. With the SRMSR, the odds to select models with more factors are smaller in large sample conditions than

in small sample conditions (i.e., 29.0 times smaller). We further note that the simple structure effect on model selection is much higher with the RMSEA and SRMSR indices than with the chi-square tests.

With WLS and WLSMV estimation, we generally see the same pattern in most selection procedures: The odds to select models with larger numbers of factors are higher with high factor loadings, simple structure, large sample size, and more response categories. With WLS, the sample size effect on model selection with RMSEA is an exception, as the odds to select models with more factors are 5.6 times smaller in large sample conditions than in small sample conditions. This might be due to the convergence problems with small sample sizes.

In the selection procedure through WLSMV estimation, the WRMSR index is new and behaves contrary to the SRMSR index in the ML procedure. For example, with the

TABLE 2
Rejection Rates According Different Estimation Methods, Various Fit Criteria and Cutoff Values, Across All Conditions With Discrete Responses

	Convergence		Chi-Square		RMSEA				SRMSR			
			($\alpha = .05$)	($\alpha = .10$)	(Cut = .03)	(Cut = .04)	(Cut = .05)	(Cut = .06)	(Cut = .02)	(Cut = .03)	(Cut = .04)	(Cut = .05)
	NC	NC/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT
<i>Series 1: Rejection rates for maximum likelihood</i>												
Sample size = 200												
Model 1	11,972	0.998	0.910	0.939	0.961	0.915	0.845	0.746	0.998	0.998	0.998	0.980
Model 2	11,385	0.949	0.701	0.750	0.809	0.730	0.640	0.536	0.949	0.949	0.907	0.655
Model 3	10,266	0.856	0.001	0.180	0.314	0.185	0.083	0.026	0.852	0.466	0.028	0.000
Model 4	4,949	0.412	0.000	0.015	0.047	0.024	0.006	0.001	0.275	0.011	0.000	0.000
Model 5	1,985	0.165	0.000	0.000	0.004	0.001	0.000	0.000	0.009	0.000	0.000	0.000
Model 6	814	0.068	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Model 7	391	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample size = 1,000												
Model 1	11,999	1.000	1.000	1.000	0.993	0.941	0.845	0.736	1.000	0.996	0.875	0.749
Model 2	11,996	1.000	0.990	0.994	0.936	0.818	0.697	0.627	0.999	0.844	0.691	0.498
Model 3	11,952	0.996	0.550	0.647	0.206	0.021	0.001	0.000	0.094	0.000	0.000	0.000
Model 4	8,666	0.722	0.137	0.204	0.036	0.002	0.000	0.000	0.000	0.000	0.000	0.000
Model 5	5,026	0.419	0.010	0.022	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Model 6	2,626	0.219	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Model 7	1,383	0.115	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>Series 2: Rejection rates for weighted least squares</i>												
Sample size = 200												
Model 1	11,791	0.983	0.973	0.977	0.980	0.974	0.959	0.921	-	-	-	-
Model 2	10,758	0.897	0.777	0.815	0.848	0.803	0.716	0.586	-	-	-	-
Model 3	9,268	0.772	0.210	0.293	0.426	0.298	0.168	0.073	-	-	-	-
Model 4	3,960	0.330	0.007	0.017	0.050	0.024	0.008	0.002	-	-	-	-
Model 5	1,394	0.116	0.000	0.000	0.002	0.001	0.000	0.000	-	-	-	-
Model 6	408	0.034	0.000	0.000	0.000	0.000	0.000	0.000	-	-	-	-
Model 7	166	0.014	0.000	0.000	0.000	0.000	0.000	0.000	-	-	-	-
Sample size = 1,000												
Model 1	12,000	1.000	1.000	1.000	0.990	0.923	0.802	0.694	-	-	-	-
Model 2	11,983	0.999	0.985	0.991	0.917	0.771	0.611	0.422	-	-	-	-
Model 3	11,914	0.993	0.514	0.621	0.156	0.010	0.000	0.000	-	-	-	-
Model 4	8,004	0.667	0.100	0.160	0.020	0.000	0.000	0.000	-	-	-	-
Model 5	4,460	0.372	0.004	0.013	0.001	0.000	0.000	0.000	-	-	-	-
Model 6	2,187	0.182	0.000	0.001	0.000	0.000	0.000	0.000	-	-	-	-
Model 7	1,285	0.107	0.000	0.000	0.000	0.000	0.000	0.000	-	-	-	-

(Continued)

TABLE 2
(Continued)

	Convergence		Chi-Square		RMSEA				WRMSR			
			($\alpha = .05$)	($\alpha = .10$)	(Cut = .03)	(Cut = .04)	(Cut = .05)	(Cut = .06)	(Cut = .40)	(Cut = .50)	(Cut = .60)	(Cut = .70)
	NC	NC/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT	NR/NT
<i>Series 3: Rejection rates for robust weighted least squares</i>												
Sample size = 200												
Model 1	11,991	0.999	0.903	0.935	0.959	0.910	0.839	0.749	0.999	0.999	0.999	0.991
Model 2	11,721	0.977	0.714	0.765	0.824	0.746	0.656	0.566	0.977	0.971	0.905	0.739
Model 3	10,899	0.908	0.058	0.115	0.246	0.119	0.038	0.007	0.653	0.219	0.021	0.001
Model 4	5,654	0.471	0.002	0.007	0.035	0.012	0.002	0.000	0.042	0.001	0.000	0.000
Model 5	2,149	0.179	0.000	0.000	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Model 6	653	0.054	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Model 7	246	0.021	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample size = 1,000												
Model 1	12,000	1.000	1.000	1.000	0.992	0.943	0.862	0.765	1.000	1.000	1.000	1.000
Model 2	11,998	1.000	0.990	0.994	0.938	0.841	0.741	0.677	1.000	1.000	1.000	0.998
Model 3	11,978	0.998	0.458	0.576	0.119	0.006	0.000	0.000	0.969	0.650	0.197	0.020
Model 4	8,989	0.749	0.102	0.164	0.018	0.000	0.000	0.000	0.297	0.031	0.001	0.000
Model 5	5,368	0.447	0.006	0.016	0.002	0.000	0.000	0.000	0.003	0.000	0.000	0.000
Model 6	2,730	0.228	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Model 7	1,946	0.162	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note. RMSEA = root mean square error of approximation; SRMSR = standardized root mean square residual; NC = number of analyses with convergence; NR = number of data sets with which the model is rejected; NT = total number of data sets; WRMSR = weighted root mean square residual. NT = 12,000 data sets with the maximum likelihood, weighted least squares, and robust weighted least squares method. Model rejection rates are calculated over 4 conditions x 3 response scale types.

WRMSR cutoff of .50, the odds to select a model of higher dimensionality in conditions with high factor loadings are 5.1 times smaller than in conditions with small factor loadings, and the odds to select a model of higher dimensionality in conditions with four-point scales are 3.7 times smaller than in conditions with dichotomous response scales.

Estimation method effects on model selection. We repeat the ordinal logistic regression analyses but include estimation method as one of the predictors of model selection. We restrict the comparison of estimation methods to the ML and WLSMV methods, as these two methods show the best convergence rates. To validly compare the two methods, we exclude all cases in which no model was selected (see Table 3), and we only consider cases in which both methods selected a model. As a result, the ordinal logistic regression analysis of model selection, reported in Table 5, has been conducted with 41,106, 45,960, and 38,070 cases for the chi-square test, the RMSEA index, and the chi-square difference test, respectively.

From Table 5, it appears that the odds to select a model with one additional factor with the ML method are 1.3 times higher than the odds with the WLSMV method, when using the chi-square test. With the chi-square difference test the ML odds are 1.2 times higher than the WLSMV odds. Strikingly, with the .05 RMSEA cut, the ML odds are 1.2 times smaller than the WLSMV odds.

All other Table 5 results are similar to the Table 4 results, with higher odds for conditions with high loadings, simple structure, large sample size, and more response categories than for the condition of reference with low loadings, complex structure, small sample size, and dichotomous responses.

Nonconvergence. All methods of EFA of discrete data seem to be plagued with convergence problems, especially when the numbers of factors increase. To compare convergence problems across the five estimation methods, Table 6 gives proportions of nonconvergence across all conditions for the three- and four-point responses that are analyzed with all estimation methods. From Table 6 it is clear that the MLR method has the highest proportions of nonconvergence and ML with polychoric correlations has the lowest. A logistic regression of nonconvergence with the four-factor model (Model 4) shows more convergence problems in conditions with low factor loadings (odds ratio [OR] = 1.2), small sample size (OR = 1.9), complex structure (OR = 1.1), three response categories (OR = 1.1), ML estimation (OR = 1.2), WLSMV estimation (OR = 1.3), ML with polychoric correlation estimation (OR = 3.2), and WLS estimation (OR = 1.1) than in the conditions with high factor loadings, large sample size, simple structure, four response categories, and MLR estimation. Analysis without the MLR methods and all categorical response scales give comparable results.

TABLE 3
 Selection Rates According to Different Estimation Methods, Various Fit Criteria and Cutoff Values. Across All Conditions With Discrete Responses

	Chi-Square			RMSEA			SRMSR			Difference Test		
	($\alpha = .05$)	($\alpha = .10$)	($\alpha = .10$)	($Cut = .03$)	($Cut = .04$)	($Cut = .06$)	($Cut = .02$)	($Cut = .03$)	($Cut = .04$)	($Cut = .05$)	($\alpha = .05$)	($\alpha = .10$)
	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT
<i>Series 1: Selection rates for maximum likelihood</i>												
Sample size = 200												
Model 1	0.087	0.017	0.036	0.083	0.153	0.252	0.000	0.000	0.000	0.017	0.000	0.000
Model 2	0.177	0.282	0.112	0.152	0.183	0.199	0.000	0.000	0.042	0.282	0.101	0.091
Model 3	0.560	0.644	0.446	0.513	0.539	0.505	0.004	0.389	0.801	0.644	0.530	0.493
Model 4	0.056	0.007	0.109	0.072	0.041	0.013	0.136	0.210	0.027	0.007	0.193	0.186
Model 5	0.009	0.003	0.024	0.015	0.006	0.002	0.111	0.031	0.007	0.003	0.016	0.016
Model 6	0.003	0.001	0.008	0.005	0.002	0.001	0.030	0.011	0.003	0.001	0.001	0.001
Model 7	0.002	0.000	0.006	0.003	0.001	0.000	0.013	0.006	0.001	0.000	0.000	0.000
No model	0.106	0.046	0.259	0.158	0.075	0.029	0.707	0.353	0.118	0.046	0.159	0.213
Sample size = 1,000												
Model 1	0.000	0.000	0.007	0.059	0.155	0.264	0.000	0.004	0.125	0.251	0.000	0.000
Model 2	0.010	0.006	0.057	0.123	0.148	0.108	0.001	0.152	0.183	0.251	0.014	0.013
Model 3	0.436	0.343	0.728	0.797	0.697	0.627	0.902	0.844	0.692	0.498	0.236	0.206
Model 4	0.286	0.287	0.130	0.017	0.000	0.000	0.068	0.000	0.000	0.000	0.339	0.307
Model 5	0.095	0.123	0.027	0.002	0.000	0.000	0.005	0.000	0.000	0.000	0.176	0.170
Model 6	0.022	0.031	0.006	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.033	0.032
Model 7	0.009	0.013	0.003	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
No model	0.142	0.197	0.042	0.002	0.000	0.000	0.021	0.000	0.000	0.000	0.202	0.272
<i>Series 2: Selection rates for weighted least squares</i>												
Sample size = 200												
Model 1	0.010	0.005	0.003	0.008	0.023	0.062	-	-	-	-	0.000	0.000
Model 2	0.113	0.078	0.047	0.087	0.162	0.260	-	-	-	-	0.041	0.030
Model 3	0.477	0.420	0.312	0.407	0.474	0.472	-	-	-	-	0.385	0.336
Model 4	0.078	0.102	0.123	0.097	0.061	0.030	-	-	-	-	0.207	0.198
Model 5	0.011	0.014	0.023	0.015	0.009	0.004	-	-	-	-	0.012	0.012
Model 6	0.002	0.003	0.005	0.003	0.001	0.001	-	-	-	-	0.000	0.000
Model 7	0.002	0.002	0.003	0.002	0.001	0.000	-	-	-	-	0.000	0.000
No model	0.309	0.375	0.484	0.381	0.268	0.171	-	-	-	-	0.355	0.424
Sample size = 1,000												
Model 1	0.000	0.000	0.010	0.077	0.198	0.306	-	-	-	-	0.000	0.000
Model 2	0.014	0.008	0.072	0.152	0.191	0.272	-	-	-	-	0.001	0.000
Model 3	0.466	0.364	0.759	0.760	0.611	0.422	-	-	-	-	0.248	0.211
Model 4	0.254	0.266	0.093	0.008	0.000	0.000	-	-	-	-	0.361	0.331
Model 5	0.068	0.095	0.016	0.001	0.000	0.000	-	-	-	-	0.149	0.144
Model 6	0.015	0.021	0.004	0.000	0.000	0.000	-	-	-	-	0.022	0.022
Model 7	0.006	0.010	0.001	0.000	0.000	0.000	-	-	-	-	0.000	0.000
No model	0.177	0.236	0.045	0.003	0.000	0.000	-	-	-	-	0.219	0.292

(Continued)

TABLE 3
(Continued)

	Chi-Square			RMSEA			WRMSR			Difference Test		
	($\alpha = .05$)	($\alpha = .10$)	($\alpha = .10$)	(Cut = .04)	(Cut = .05)	(Cut = .06)	(Cut = .40)	(Cut = .50)	(Cut = .60)	(Cut = .70)	($\alpha = .05$)	($\alpha = .10$)
	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT	NS/NT
<i>Series 3: Selection rates for robust weighted least squares</i>												
Sample size = 200												
Model 1	0.097	0.065	0.040	0.089	0.160	0.250	0.000	0.000	0.000	0.009	0.000	0.000
Model 2	0.176	0.154	0.118	0.151	0.175	0.180	0.000	0.006	0.071	0.231	0.111	0.104
Model 3	0.644	0.633	0.550	0.613	0.610	0.555	0.255	0.684	0.832	0.719	0.638	0.612
Model 4	0.034	0.060	0.103	0.056	0.022	0.006	0.282	0.109	0.018	0.006	0.154	0.152
Model 5	0.004	0.008	0.020	0.009	0.003	0.001	0.043	0.013	0.004	0.001	0.008	0.008
Model 6	0.001	0.002	0.004	0.002	0.001	0.000	0.009	0.003	0.001	0.000	0.000	0.000
Model 7	0.000	0.001	0.002	0.001	0.000	0.000	0.003	0.001	0.000	0.000	0.000	0.000
No model	0.044	0.077	0.163	0.079	0.028	0.009	0.408	0.184	0.073	0.034	0.089	0.124
Sample size = 1,000												
Model 1	0.000	0.000	0.008	0.057	0.138	0.236	0.000	0.000	0.000	0.000	0.000	0.000
Model 2	0.010	0.006	0.054	0.103	0.121	0.088	0.000	0.000	0.000	0.002	0.001	0.001
Model 3	0.531	0.417	0.819	0.835	0.741	0.677	0.029	0.348	0.801	0.976	0.287	0.253
Model 4	0.269	0.294	0.084	0.005	0.000	0.000	0.429	0.447	0.140	0.015	0.379	0.353
Model 5	0.076	0.107	0.015	0.000	0.000	0.000	0.181	0.043	0.011	0.001	0.172	0.168
Model 6	0.016	0.025	0.003	0.000	0.000	0.000	0.033	0.012	0.004	0.000	0.030	0.029
Model 7	0.007	0.012	0.002	0.000	0.000	0.000	0.023	0.008	0.002	0.000	0.000	0.000
No model	0.091	0.138	0.016	0.000	0.000	0.000	0.304	0.142	0.043	0.006	0.131	0.197

Note. RMSEA = root mean square error of approximation; SRMSR = standardized root mean square residual; NS = number of data sets with which the model is selected; NT = total number of data sets; WRMSR = weighted root mean square residual. NT contains 12,000 data sets with the maximum likelihood, weighted least squares, and robust weighted least squares method. Model selection rates are calculated over 4 conditions \times 3 response scale types.

TABLE 4
Ordinal Logistic Regression Analysis of Model Selection, for Different Estimation Methods According to Various Fit Criteria, Across All Conditions With Discrete Responses

Conditions	Chi-Square ($\alpha = .05$)		RMSEA (Cut = .05)		SRMSR (Cut = .04)		Difference Test ($\alpha = .05$)					
	β	OR	β	OR	β	OR	β	OR				
<i>Ordinal logistic regression analysis of model selection through maximum likelihood estimation</i>												
High loadings	2.513	12.341	5.372	215.202	3.795	44.456	2.199	9.019				
Simple structure	0.562	1.755	2.786	16.211	3.465	31.968	0.004	1.004				
Large sample size	3.746	42.339	0.102	1.107	-3.366	0.035	2.773	16.009				
3 response categories	1.158	3.185	1.584	4.873	1.427	4.167	2.217	9.181				
4 response categories	1.590	4.903	2.135	8.459	1.694	5.441	1.067	2.906				
Continuous responses	2.775	16.038	3.365	28.934	1.968	7.159	1.473	4.360				
<i>Ordinal logistic regression analysis of model selection through weighted least squares estimation</i>												
High loadings	1.566	4.787	4.037	56.639	-	-	1.175	3.240				
Simple structure	0.634	1.886	3.006	20.201	-	-	0.546	1.726				
Large sample size	2.226	9.261	-1.729	0.177	-	-	1.993	7.336				
3 response categories	1.249	3.488	1.604	4.973	-	-	1.156	3.178				
4 response categories	1.816	6.145	2.205	9.073	-	-	1.621	5.059				
<i>WRMSR (Cut = .50)</i>												
<table border="0" style="width: 100%;"> <tr> <td style="width: 50%;"></td> <td style="width: 50%;"></td> </tr> <tr> <td style="text-align: center;">β</td> <td style="text-align: center;">OR</td> </tr> </table>											β	OR
β	OR											
<i>Ordinal logistic regression analysis of model selection through robust weighted least squares estimation</i>												
High loadings	2.287	9.845	1.959	7.092	-1.633	0.195	1.401	4.060				
Simple structure	1.002	2.724	0.832	2.298	-0.032	0.969	0.570	1.769				
Large sample size	3.722	41.341	3.388	29.615	2.304	10.009	2.961	19.313				
3 response categories	1.362	3.902	1.252	3.498	-0.694	0.499	1.112	3.042				
4 response categories	1.937	6.938	1.814	6.134	-1.318	0.268	1.647	5.191				

Note. RMSEA = root mean square error of approximation; SRMSR = standardized root mean square residual; OR = odds ratio; WRMSR = weighted root mean square residual. The condition of reference has small factor loadings, complex structure, small sample size, and dichotomous response scales. With maximum likelihood estimation, the regression analysis is based on 32,000 data sets and with weighted least squares and robust weighted least squares on 24,000 data sets.

TABLE 5
Ordinal Logistic Regression of Model Selection With the Maximum Likelihood and Robust Weighted Least Squares Methods, Across All Conditions With Dichotomous, Three-Point, and Four-Point Response Scales

Conditions	Chi-Square ($\alpha = .05$)		RMSEA (Cut = .05)		Difference Test ($\alpha = .05$)	
	β	OR	β	OR	β	OR
Maximum likelihood estimation	0.291	1.338	-0.167	0.846	0.182	1.199
High loadings	2.476	11.895	6.178	481.942	1.695	5.447
Simple data structure	0.799	2.224	3.747	42.375	0.343	1.410
Large sample size	3.656	38.725	0.464	1.590	2.816	16.716
3 response categories	1.294	3.649	1.777	5.912	1.086	2.962
4 response categories	1.819	6.167	2.434	11.404	1.559	4.753

Note. RMSEA = root mean square error of approximation; OR = odds ratio. β is the log odds ratio and OR is the ratio of odds of model selection in the particular condition over the odds in the condition of reference. The condition of reference is small loadings, complex structure, small sample size, and dichotomous response scales, analyzed with the weighted least squares method. For chi-square tests, odds ratios are calculated for 41,106 data sets: 13,796, 13,682, and 13,628 cases with dichotomous, three-point, and four-point responses, respectively. For RMSEA indices, odds ratios are calculated for 45,960 data sets: 15,290, 15,342, and 15,328 cases with dichotomous, three-point, and four-point responses, respectively. For chi-square difference tests, odds ratios are calculated for 38,070 data sets: 12,550, 12,810, and 12,710 cases with dichotomous, three-point, and four-point responses, respectively.

DISCUSSION

We compared different estimation methods, with different fit criteria, under various conditions, but judgment of preferred methods and criteria depends on the goals of the

user. We used a model with three major factors and four minor factors to generate data with six dimensions (six, not seven, because of our parameterization choices). If the researcher would like to recover six factors, he or she might prefer the chi-square test that is a measure of exact fit.

TABLE 6
Proportions of Nonconvergence, Calculated Across All Conditions
With Three-Point and Four-Point Response Scales

	<i>Estimation Methods</i>				
	<i>ML</i>	<i>MLR</i>	<i>WLSMV</i>	<i>Polychoric ML</i>	<i>WLS</i>
Model 1	0.000	0.001	0.000	0.001	0.002
Model 2	0.018	0.048	0.005	0.012	0.027
Model 3	0.057	0.118	0.027	0.033	0.074
Model 4	0.408	0.462	0.331	0.214	0.439
Model 5	0.681	0.731	0.634	0.328	0.711
Model 6	0.840	0.868	0.826	0.379	0.864
Model 7	0.922	0.938	0.890	0.503	0.928

Note. ML = maximum likelihood; MLR = robust maximum likelihood; WLSMV = robust weighted least squares; WLS = weighted least squares. The total number of data sets in these conditions is 16,000.

The chi-square test and chi-square difference test procedures generally select models with more common factors than the RMSEA and the SRMSR or WRMSR procedures, but hardly ever the six-factor model. This is consistent with the study of Briggs and MacCallum (2003) with continuous data that showed difficulties recovering patterns with minor factors that were even larger than ours. In contrast, simulation studies of models with only major factors encountered overextraction problems (e.g., Beauducél, 2001; Hayashi et al., 2007). The chi-square difference test procedure recovers more factors than the stand-alone chi-square test procedure, but it also more frequently ends up without selecting any model, because of nonconvergence problems.

If the researcher is primarily interested in major factors, he or she might want to rely on the RMSEA index. For example, with the robust WLS estimation method, and with the common choice of a .05 cutoff, the RMSEA selects the three-factor model in 61.0% of the small sample cases and in 74.1% of the large sample cases. For the chi-square test procedure, these selection percentages are 64.4% and 53.1%. Interestingly, the WRMSR with .60 and .70 cutoffs outperformed the RMSEA procedure in selecting the three-factor model. Obviously, selection rates largely depend on the (arbitrary) choices of a level of significance for the chi-square tests and cutoffs for the RMSEA and RMSR indices. Marsh, Hau, and Wen (2004) and Hooper, Coughlan, and Mullen (2008) discussed model selection guidelines for model selection in structural equation modeling of variance and covariances of continuous data; less is known about discrete data.

Selection rates also depend on circumstances. Chi-square tests and RMSEA index procedures recover more factors when sample size is larger, when factor loadings are higher, when factor structure is less complex, and when the response scales have more options. The effects of sample size and communality (size of factor loadings) have also been found in studies of factor pattern recovery (MacCallum, Widaman, Preacher, & Hong, 2001; MacCallum, Widaman, Zhang, & Hong, 1999). When model selection is based on the SRMSR

index, we find an opposite effect of sample size, as the SRMSR gets smaller with larger sample size, which can be explained by the more precise correlation estimates on which the SRMSR is based. Conspicuously, all effects on model selection with the WRMSR index are opposite to those of the SRMSR index.

For the purpose of clarity of presentation, we have only reported main effects of sample size, factor loading size, factor structure complexity, and number of response options. However, for example, as apparent in Table 1, the effect of the factor loading size varies with sample size and factor structure complexity. Full results of all estimation methods and fit criteria can be downloaded from the website of the corresponding author.

With all estimation methods, EFA is hindered by convergence problems, especially when the numbers of factors increase and when the sample size is small. To prevent convergence problems, we examined various rotation constraints and we experimented with starting values. With the *Mplus* computer program, the echelon form rotation constraint appears to work best. Of all estimation methods that have been used in our study, the WLSMV method is the preferred method of choice. The ML methods lack theoretical justification for use with discrete data. The MLR method has the problem that the chi-square difference test often yields negative results, although we should note that Satorra and Bentler recently suggested a new procedure that consistently yields positive results (Asparouhov & Muthén, 2010; Satorra & Bentler, 2010). In our study, the ML estimation and WLSMV estimation do not show large differences in model selection, but Beauducél and Herzberg (2006) showed that the WLSMV estimates are more precise with only a few response options.

In conclusion, the WLSMV estimation method is theoretically justified for the factor analysis of discrete data and it gives the best results. In addition, this limited information method of analyzing polychoric correlations is more practical, as full information methods such as maximizing the likelihood of observed response patterns are too computationally intensive to be used in practice, even with small numbers of variables (Jöreskog & Moustaki, 2001; Wirth & Edwards, 2007).

FUNDING

This publication is supported by an open competition grant 400-09-084 from the Netherlands Organization for Scientific Research (NWO).

REFERENCES

- Asparouhov, T., & Muthén, B. (2010). Simple second order chi-square correction. *Mplus Technical Appendix*. Retrieved from http://statmodel.com/download/WLSMV_new_chi21.pdf

- Asparouhov, T., Muthén, M., & Muthén, B. (2006). Robust chi square difference testing with mean and variance adjusted test statistics. *Mplus Web Notes*. Retrieved from <http://www.statmodel.com/download/webnotes/webnote.pdf>
- Beauducel, A. (2001). On the generalizability of factors: The influence of changing contexts of variables on different methods of factor extraction. *Methods of Psychological Research Online*, 6, 1–28.
- 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.
- Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, 38, 25–56.
- Browne, M. W. (1982). Covariance structures. In *Topics in applied multivariate analysis*, ed. D. M. Hawkins (pp. 72–141). Cambridge, UK: Cambridge University Press.
- Browne, M. W. (1984). Asymptotically distribution-free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, 21, 230–258.
- 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.
- Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, 47, 309–326.
- 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.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9, 466–491.
- Hayashi, K., Bentler, P. M., & Yuan, K. H. (2007). On the likelihood ratio test for the number of factors in exploratory factor analysis. *Structural Equation Modeling*, 14, 505–526.
- Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity*, 44, 153–166.
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Journal of Business Research Methods*, 6, 53–60.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, 48, 398–407.
- Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36, 347–387.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149.
- MacCallum, R. C., Widaman, K. F., Preacher, K., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, 36, 611–637.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84–99.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modeling*, 11, 320–341.
- Muthén, B. O. (1998–2004). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén. Retrieved from <http://statmodel.com/download/techappen.pdf>
- Muthén, B., du Toit, S. H., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika*, 75, 1–45.
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171–189.
- Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45, 19–30.
- Muthén, B. O., & Muthén, L. K. (2010). *Mplus user's guide: Statistical analysis with latent variables*. Los Angeles, CA: Muthén & Muthén.
- Olsson, U. H., Foss, T., & Breivik, E. (2004). Two equivalent discrepancy functions for maximum likelihood estimation: Do their test statistics follow a non-central chi-square distribution under model misspecification? *Sociological Methods Research*, 32, 453–500.
- Olsson, U. H., Troye, S. V., & Howell, R. D. (1999). Theoretic fit and empirical fit: The performance of maximum likelihood versus generalized least squares estimation in structural equation modeling. *Multivariate Behavioral Research*, 34, 31–58.
- Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2, 13–32.
- Rhemtulla, M., Brosseau-Liard, P., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354.
- Rigdon, E. E., & Ferguson, C. E. (1991). The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with discrete data. *Journal of Marketing Research*, 28, 491–497.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507–514.
- Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, 75, 243–248.
- 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.
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50, 253–264.
- Ten Holt, J. C., van Duijn, M. A. J., & Boomsma, A. (2010). Scale construction and evaluation in practice: A review of factor analysis versus item response theory. *Psychological Test and Assessment Modeling*, 52, 272–297.
- Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, 12, 58–79.
- Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, 17, 392–423.

- Yu, C. Y., & Muthén, B. (2002). *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes* (Technical report). Los Angeles, CA: University of California at Los Angeles, Graduate School of Education & Information Studies.
- Yuan, K. H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modelling. *British Journal of Mathematical and Statistical Psychology*, *51*, 289–309.
- Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In M. E. Sobel & M. P. Becker (Eds.), *Sociological methodology 2000* (pp. 165–200). Washington, DC: ASA.
- Zhang, J., & Stout, W. F. (1999). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, *64*, 129–152.