

**On the exploratory road to unravelling factor loading non-invariance: A new  
multigroup rotation approach**

Kim De Roover

Tilburg University, KU Leuven

Jeroen K. Vermunt

Tilburg University

Author Notes:

The research leading to the results reported in this paper was funded by the Netherlands Organization for Scientific Research (NWO) [Veni grant 451-16-004]. Correspondence concerning this paper should be addressed to Kim De Roover, Tilburg School of Social and Behavioral Sciences, Department of Methodology and Statistics, PO box 90153 5000 LE Tilburg, The Netherlands. E-mail: K.DeRoover@uvt.nl.

### **Abstract**

Multigroup exploratory factor analysis (EFA) has gained popularity to address measurement invariance for two reasons. Firstly, repeatedly respecifying confirmatory factor analysis (CFA) models strongly capitalizes on chance and using EFA as a precursor works better. Secondly, fixed zero loadings are often too restrictive. In multigroup EFA, factor loading invariance is untenable if the fit decreases significantly when fixing the loadings to be equal across groups. To locate the precise factor loading non-invariances by means of hypothesis testing, the factors' rotational freedom needs to be resolved per group. In the literature, a solution exists for identifying optimal rotations for one group or invariant loadings across groups. Building on this, we present multigroup factor rotation (MGFR) for identifying loading non-invariances. Specifically, MGFR rotates group-specific loadings both to simple structure and between-group agreement, while disentangling loading differences from differences in the structural model (i.e., factor (co)variances).

*Keywords: measurement invariance; factor loading invariance; multigroup exploratory factor analysis; rotation identification*

## 1. Introduction

In behavioral sciences, latent constructs, e.g., emotions or personality traits, are ubiquitously measured by questionnaire items. The measurement model (MM) indicates which item is (assumed to be) measuring which construct and the leading method to evaluate whether this MM holds is confirmatory factor analysis (CFA; Lawley & Maxwell, 1962). The extent to which an item relates to a construct or ‘factor’ is quantified by a ‘factor loading’. CFA evaluates whether each item has a non-zero loading on the targeted construct only. Many research questions pertain to comparing constructs across groups, e.g., comparing the Big Five personality traits across countries (Schmitt, Allik, McCrae, & Benet-Martinez, 2007). For such comparisons, invariance of the MM or ‘measurement invariance’ (MI) across the groups is an essential prerequisite (Meredith, 1993). MI can be tested by multigroup factor analysis (Jöreskog, 1971; Sörbom, 1974). Despite the predominance of CFA-based methods, multigroup *exploratory* factor analysis (EFA) has gained popularity to address MI (Dolan, Oort, Stoel, & Wicherts, 2009; Marsh, Morin, Parker, & Kaur, 2014). The reason for this is twofold. Firstly, respecifying CFA models in an exploratory way capitalizes on chance (Browne, 2001; MacCallum, Roznowski, & Necowitz, 1992) and using EFA as a precursor has proven to be a better strategy (Gerbing & Hamilton, 1996). Secondly, fixed zero loadings are often too restrictive and may cause bias (Muthén & Asparouhov, 2012; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996).

MI testing with multigroup EFA starts by evaluating whether the fit significantly decreases when fixing the factor loadings to be equal (i.e., invariant) across groups, indicating that factor loading (or ‘weak’) invariance does not hold. Because EFA is used within the groups, the factors have rotational freedom, i.e., ‘rotating’ them yields an alternative set of factors which fit equally well to the data but may be easier to interpret (Brown, 2001; Osborne, 2015). When merely testing

invariance for all loadings, the factor rotation is irrelevant. However, the rotation becomes of interest when one wants to determine what the invariant MM is (Asparouhov & Muthén, 2009; Dolan et al., 2009). To this end, simple structure rotation (i.e., striving for one non-zero loading per item; Thurstone, 1947) or target rotation towards an assumed MM can be applied. To enable hypothesis testing for rotated loadings, Jennrich (1973) showed how to obtain a fully identified model with optimally rotated maximum likelihood (ML) estimates.

Jennrich's approach does the trick for single-group factor models and multigroup factor models with invariant loadings, but leaves much to be desired when loadings are non-invariant across groups. In that case, pinpointing the precise loading differences would allow to find sources of non-invariance and interesting differences in the functioning of items (differential item functioning or DIF; Holland & Wainer, 1993). To this end, an optimal rotation needs to be obtained for each group. Using Jennrich's approach per group precludes pursuing optimal between-group agreement of the loadings and thus impedes a correct evaluation of differences and similarities. Therefore, we present a multigroup extension to accommodate the search for loading differences. Specifically, each group is rotated both to simple structure per group and agreement across groups and, at the same time, loading differences are disentangled from differences irrelevant to the MI question (i.e., factor (co)variances). The novel multigroup factor rotation (MGFR) can be applied with several rotation criteria and with a user-specified focus on agreement or simple structure.

The remainder of this paper is organized as follows: Section 2 recaps MI testing by multigroup EFA, followed by a discussion of optimal rotation identification including the novel MGFR. Section 3 covers an extensive simulation study to evaluate the performance of MGFR with regard to the identification of loading differences and group-specific MMs and derives

recommendations for empirical practice. Section 4 illustrates the added value of MGFR for an empirical data set. Section 5 includes points of discussion and directions for future research.

## 2. Method

### 2.1. Multigroup exploratory factor analysis

We denote the groups by  $i = 1, \dots, I$  and the subjects within the groups by  $n_i = 1, \dots, N_i$ . The  $J$ -dimensional random vector of observed item scores for subject  $n_i$  is denoted by  $\mathbf{y}_{n_i}$ . The EFA model for the observations of subject  $n_i$  can be written as (Lawley & Maxwell, 1962):

$$\mathbf{y}_{n_i} = \boldsymbol{\tau}_i + \boldsymbol{\Lambda}_i \boldsymbol{\eta}_{n_i} + \boldsymbol{\varepsilon}_{n_i} \quad (1)$$

where  $\boldsymbol{\tau}_i$  indicates a  $J$ -dimensional group-specific intercept vector,  $\boldsymbol{\Lambda}_i$  denotes a  $J \times Q$  matrix of group-specific factor loadings,  $\boldsymbol{\eta}_{n_i}$  is a  $Q$ -dimensional vector of scores on the  $Q$  factors and  $\boldsymbol{\varepsilon}_{n_i}$  is a  $J$ -dimensional vector of residuals. The factor scores are assumed to be identically and independently distributed (i.i.d.) as  $MVN(\boldsymbol{\alpha}_i, \boldsymbol{\Psi}_i)$ , independently of  $\boldsymbol{\varepsilon}_{n_i}$ , which are i.i.d. as  $MVN(\mathbf{0}, \mathbf{D}_i)$ . The factor means of group  $i$  are denoted by  $\boldsymbol{\alpha}_i$ , whereas  $\boldsymbol{\Psi}_i$  pertains to the group-specific factor covariance matrix and  $\mathbf{D}_i$  to a diagonal matrix containing the group-specific unique variances of the items. The model-implied covariance matrix for group  $i$  is  $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i \boldsymbol{\Psi}_i \boldsymbol{\Lambda}_i' + \mathbf{D}_i$ .

Estimating Equation 1 for each group corresponds to the baseline model for MI testing. To partially identify the model, the factor means  $\boldsymbol{\alpha}_i$  are fixed to zero and the factor covariance matrix  $\boldsymbol{\Psi}_i$  to identity (i.e., orthonormal factors) per group  $i$ . Note that, unlike multigroup EFA, multigroup CFA imposes zero loadings on  $\boldsymbol{\Lambda}_i$  according to an assumed MM and it assumes this pattern of zero loadings to be invariant across groups (configural invariance; Meredith, 1993).

To test for MI, a series of progressively more restricted models is fitted. Factor loading invariance is evaluated by comparing the fit of the baseline model and the model with invariant loadings, i.e.,  $\Lambda_i = \Lambda$  for  $i = 1, \dots, I$ . For the latter model, orthonormality of the factors is no longer imposed per group but, e.g., for the mean factor (co)variances across groups. In the literature, several criteria and guidelines are discussed to evaluate whether a drop in fit is statistically significant (Hu & Bentler, 1999). When it is not significant, factor loading (or weak) invariance is established and the next level of MI – which is beyond the scope of this paper – can be tested by restricting the intercepts  $\tau_i$  to be common across groups, while freely estimating factor means  $\alpha_i$  per group (Dolan et al., 2009; Meredith, 1993). When the fit is significantly worse with invariant factor loadings – i.e., factor loading invariance is rejected – one can scrutinize the baseline model to locate factor loading non-invariances.

Note that, in case of multigroup CFA, the baseline model is already very restrictive due to the assumption of configural invariance. Therefore, multigroup CFA extensions for dealing with loading non-invariances – such as multigroup Bayesian structural equation modeling (multigroup BSEM; Muthén & Asparouhov, 2013) and multigroup factor alignment (Asparouhov & Muthén, 2014) – only capture differences in the size of primary loadings, whereas differences in crossloadings and the position of primary loadings are disregarded.

Thus, multigroup EFA has the important advantage that it leaves room to evaluate (the lack of) MI without having to predefine the MM and to find all types and combinations of loading differences. In the baseline model (Equation 1), the rotational freedom of the factors per group is beneficial to this aim. Specifically, striving for simple structure per group (e.g., Clarkson & Jennrich, 1988) as well as between-group agreement (e.g., ten Berge, 1977) allows for the group-

specific MMs to be identified and loading differences to be pinpointed. Thus, sources of configural and weak non-invariance can be traced simultaneously.

Multigroup EFA can be estimated by open-source software such as Lavaan (Rosseel, 2012) and Mx (Neale, Boker, Xie, & Maes, 2003) as well as commercial software such as Latent Gold (LG; Vermunt & Magidson, 2013) and Mplus (Muthén & Muthén, 2005). LG-syntax for multigroup EFA with (optimally rotated) group-specific loadings is given in Appendix A.

## ***2.2. Optimal rotation in multigroup EFA***

### *2.2.1. In case of factor loading invariance*

To partly identify a single EFA solution, up to rotation,  $Q(Q+1)/2$  restrictions are needed. Usually, the factor covariance matrix is restricted to be an identity matrix. In case of multigroup EFA with invariant loadings  $\Lambda$ , these restrictions are imposed across groups, e.g., for the mean factor (co)variances, or for one ‘reference’ group (Hessen, Dolan & Wicherts, 2006). To obtain a fully identified model, i.e., with an identified rotation, a total of  $Q^2$  restrictions are necessary, yet not always sufficient (Jöreskog, 1979). Jennrich (1973) derived the necessary restrictions for obtaining the optimal rotation according to a criterion of choice. This solution can be readily applied to rotate invariant loadings in multigroup EFA. In this paper, we focus on oblique rotation, which implies that factor correlations are not fixed to zero and that  $Q(Q-1)$  additional restrictions are needed. Specifically, to obtain an optimal oblique rotation according to rotation criterion  $R$ , the following matrix  $F$  is restricted to be diagonal:

$$F = \Lambda' \frac{dR}{d\Lambda} \Psi^{-1} . \quad (2)$$

Imposing this by constrained ML estimation is prone to convergence problems and local optima in complex cases (Asparouhov & Muthén, 2009). Jennrich (2001, 2002) proposed the

gradient projection algorithm as an alternative for imposing the constraints. Upon identifying the rotation, standard errors (SEs) for the model parameters and hypothesis testing to determine significant factor loadings, and thus the invariant MM, are available (Jennrich, 1973).

### 2.2.1.1. Simple structure rotation criteria

For the choice of rotation criterion  $R$ , several simple structure (SS) rotation criteria exist that minimize either the variable complexity (i.e., the number of non-zero loadings per variable), factor complexity (i.e., the number of non-zero loadings per factor), or a combination of both (Schmitt & Sass, 2011). We focus on oblique simple structure rotation to minimize the variable complexity since this matches the concept of a MM, i.e., items as pure measurements of one factor. Geomin (Yates, 1987) is a popular criterion (e.g., it is default in Mplus; Asparouhov & Muthén, 2009) but is sensitive to local minima (Asparouhov & Muthén, 2009; Browne, 2001). (Direct) oblimin<sup>1</sup> (Clarkson & Jennrich, 1988) is a widely-used rotation offered in the statistical packages SPSS (Nie, Bent, & Hull, 1970) and STATA (Hamilton, 2012). Stepwise rotation procedures such as promax (Hendrickson, & White, 1964) and promin (Lorenzo-Seva, 1999) cannot be readily applied as the rotation criterion in Equation 2. Simple structure rotation criteria often perform suboptimal when the variable complexity is higher than one for some items (Ferrando & Lorenzo-Seva, 2000; Lorenzo-Seva, 1999; Schmitt & Sass, 2011). To avoid this deficiency, weighted oblimin (Lorenzo-Seva, 2000) aims to downweight the influence of variables with a higher complexity in the rotation, but the weighting procedure is known to fail in some cases (Kiers, 1994). Anyway, target rotation (Browne, 2001) towards a zero-loading-pattern is a good way to achieve simple structure, since crossloadings can be tolerated by leaving the corresponding element of the target unspecified. Simplimax (Kiers, 1994) can be used to determine the optimal

---

<sup>1</sup> Oblimin performs best when the parameter  $\gamma$  is equal to zero (Jennrich, 1979) and then it is in fact direct quartimin rotation (Jennrich & Sampson, 1966), but we will refer to it as ‘oblimin’ throughout the rest of the paper.



target for a given loading matrix. When one has prior beliefs about the MM, a target corresponding to this MM can be applied. In this paper, the oblimin criterion is applied as  $R^{SS}$ , where  $\lambda_{jq}$  is the loading of item  $j$  on factor  $q$ :

$$R^{SS} = \sum_{q=1}^Q \sum_{q'=q+1}^Q \sum_{j=1}^J \lambda_{jq}^2 \lambda_{jq'}^2. \quad (3)$$

### 2.2.2. In case of factor loading non-invariance

If invariant factor loadings are untenable, the group-specific loadings are scrutinized to identify sources of non-invariance. To this end, the optimal rotation needs to be identified for each group and one may choose to apply the restrictions in Equation 2 to each group separately, implying  $I \times Q(Q-1)$  restrictions, while the factor variances remain fixed to one per group (Section 2.1). This approach entails two pitfalls. Firstly, the rotation for each group separately disregards the resulting (dis)agreement of loadings across groups, resulting in overestimated differences. Secondly, when keeping the factor variances fixed to one per group during the rotation, differences in factor scale show up in the loadings, while these differences are irrelevant to the MI question. Specifically, factor variances (as well as factor covariances) are part of the structural model rather than the MM (Dolan et al., 2009; Meredith, 1993).

To strive for agreement and simple structure, MGFR applies multigroup criterion  $R_{mg}$ :

$$R_{mg}(\Lambda_1, \dots, \Lambda_i, \dots, \Lambda_I) = w R^A + (1-w) \sum_{i=1}^I R_i^{SS} \quad (4)$$

where  $R^A$  refers to the agreement criterion across all groups and  $R_i^{SS}$  refers to a simple structure criterion within group  $i$ . For  $R^A$ , we consider two criteria discussed in Section 2.3.2.1. For  $R_i^{SS}$ , a simple structure criterion of choice may be applied (oblimin, geomin and target rotation are currently supported; see Appendix A). The relative influence of the agreement and simple

structures on  $R_{mg}$  is determined by the weighting parameter  $w$ . Thus, the novelty of this criterion lies not only in combining  $R^A$  and  $R_i^{SS}$  ( $i = 1, \dots, I$ )<sup>2</sup> but also in the weighting of this combination, resulting in a flexible framework of rotations that includes every degree of focus on either agreement or simple structure. Note that adding  $R_i^{SS}$  to  $R^A$  not only allows to evaluate the MMs of the groups but also optimizes the power for detecting loading differences (Pennell, 1968).

To identify the scales of the group-specific factors, we restrict the mean factor variances to one:  $\frac{1}{I} \sum_{i=1}^I \text{diag}(\Psi_i) = [\mathbf{1}]_Q$ . As such, we allow for factor variances to differ between groups and avoid the arbitrariness of choosing a reference group with fixed variances. The group-specific factor variances will be further identified by the  $R^A$  part (i.e., maximizing between-group agreement), whereas the factor covariances are identified by both parts of the optimal rotation (i.e., maximizing simple structure per group as well as between-group agreement).

Given the  $Q$  scaling restrictions,  $(I-1)Q^2 + Q(Q-1)$  additional restrictions are needed to identify the optimal multigroup rotation. To find the restrictions that minimize  $R_{mg}$ , we use the differential of  $R_{mg}$  in the point corresponding to the optimally rotated loadings  $\Lambda_i$  for  $i = 1, \dots, I$ :

$$dR_{mg}(\Lambda_1, \dots, \Lambda_i, \dots, \Lambda_I) = \sum_{i=1}^I dR_{mg}(\Lambda_i) \quad (5)$$

The differential is derived in Appendix B and results in the following restrictions for each group:

$$F_{mg,i} = \Lambda_i' \frac{dR_{mg}}{d\Lambda_i} \Psi_i^{-1} - \text{diag} \left( \frac{1}{I} \sum_{i'=1}^I \Lambda_{i'}' \frac{dR_{mg}}{d\Lambda_{i'}} \Psi_{i'}^{-1} \right) = [\mathbf{0}]_{Q \times Q} \quad (6)$$

---

<sup>2</sup> Note that Lorenzo-Seva, Kiers and ten Berge (2002) already presented a set of oblique rotations of multiple loading matrices to a compromise of simple structure and optimal agreement. These rotations are performed in a stepwise manner, however, making them hard to implement in a single rotation criterion in MGFR. Also, they either do not allow for differences in factor correlations between the groups or do not maintain between-group agreement in the final step, resulting in a suboptimal between-group agreement of the rotated loadings.

Again, SEs can be obtained for the optimally rotated loadings (Jennrich, 1973) and hypothesis testing can be performed. To identify for factor loading non-invariances, one can test per loading whether it is significantly different across the groups using a Wald test. When one is interested in MMs (or causes of configural non-invariance), one can also test which loadings are significantly different from zero per group and evaluate how these results differ across groups.

### 2.2.2.1. Agreement rotation criteria

A widely used criterion for agreement rotation of multiple loading matrices is generalized procrustes (GP; ten Berge, 1977), which optimizes agreement in the least squares sense:

$$R^A = \sum_{i=1}^I \sum_{i'=i+1}^I \sum_{j=1}^J \sum_{q=1}^Q (\lambda_{ijq} - \lambda_{i'jq})^2 \quad (7)$$

Due to the square, the loss due to a loading difference smaller than one is attenuated, and more so for smaller differences. The loss due to a difference larger than one is amplified. Thus, GP aims to minimize large loading differences and tolerate small differences. This implies the disadvantage that, in the attempt to minimize (true) large differences, (false) small differences may be created. Note that GP is originally an orthogonal rotation, but in line with the oblique simple structure rotations, we will not impose orthogonality when using GP in MGFR. Thus, as an added value over GP, MGFR not only offers the combination with  $R_i^{SS}$  but also the disentanglement of loading differences from differences in factor variances and correlations.

As an alternative, some aspects of the (confirmatory) multigroup factor alignment (Asparouhov & Muthén, 2014) can be included in MGFR. Specifically, in multigroup factor alignment, the factors are ‘aligned’ (i.e., rescaled and shifted in terms of their factor means) to minimize the following function of loading and intercept differences, separately per factor  $q$ :

$$\sum_{i=1}^I \sum_{i'=i+1}^I \sum_{j=1}^J w_{ii'} \left( \sqrt{\sqrt{(\lambda_{ijq} - \lambda_{i'jq})^2} + \varepsilon} + \sqrt{\sqrt{(\tau_{ij} - \tau_{i'j})^2} + \varepsilon} \right) \quad (8)$$

where  $\varepsilon$  is a small number included to facilitate the minimization and  $w_{ii'} = \sqrt{N_i N_{i'}}$  is a weight depending on the group sizes. On the one hand, intercept (and factor mean) differences are beyond the scope of this paper and are omitted from the criterion (i.e., they are fixed during rotation) for MGFR. On the other hand, we are dealing with (the rotation of) EFA rather than CFA and thus apply the criterion across all factors simultaneously. Therefore, it becomes:

$$R^A = \sum_{i=1}^I \sum_{i'=i+1}^I \sum_{j=1}^J \sum_{q=1}^Q \sqrt{\sqrt{(\lambda_{ijq} - \lambda_{i'jq})^2 + \varepsilon}} \quad (9)$$

where  $w_{ii'}$  is omitted since  $R_i^{SS}$  does not include such a weight. We will refer to this adjusted alignment criterion as the ‘loading alignment’ (LA) criterion. The square root attenuates the loss for loading differences larger than one, whereas the loss is amplified for differences smaller than one, and more so for small differences. Therefore, minimizing the LA criterion eliminates small loading differences while large differences are tolerated. Thus, it strives for loading differences to be either zero or large (Asparouhov & Muthén, 2014), which fits our aim of distinguishing invariant from non-invariant loadings irrespective of the size of the non-invariance.

Using the LA criterion as  $R^A$  in MGFR provides an important added value to multigroup factor alignment. Firstly, the criterion is applied not only to rescale factors but also to (obliquely) rotate them towards agreement. Secondly, the oblique rotation implies that differences in factor loadings are now disentangled not only from differences in factor variances but also from differences in factor correlations. Thirdly, the factors are dealt with simultaneously. Fourthly, it is combined with simple structure rotation per group. Of course, there is more to multigroup factor alignment than Equation 8. In Section 5, we discuss aspects that are not included in MGFR, as well as the relevance and possibilities to add them in the future.

Note that the combination with  $R_i^{SS}$ , for  $i = 1, \dots, I$ , implies the following: (1) The value of  $\varepsilon$  in Equation 9 affects the relative weight of  $wR^A$  over the simple structure part, and more so in case of a higher  $I$ ,  $J$  and/or  $Q$ . To minimize this effect, the value of  $\varepsilon$  should be kept very small (e.g.,  $1 \times 10^{-12}$ ), or its influence compensated for by decreasing  $w$ . (2) Given a certain  $w$  and the size of loading differences, the weight of  $wR^A$  will strongly depend on whether the agreement is pursued by GP or LA. Thus, the fine-tuning of  $w$  will depend on the  $R^A$  that is used.

### 2.2.3. Implementation of optimal rotation

MGFR is implemented in LG 6.0 and applied by syntax (Appendix A). In the future, it can be readily implemented in other software (e.g., Lavaan, Mplus). The performed steps are:

1. *ML estimation*: The model is estimated without the optimal rotation restrictions, i.e., maximizing the log-likelihood ( $LL$ ), with factor variances fixed to one per group.
2. *Gradient projection per group*: Using the estimates from Step 1 as initial values and keeping the factor variances fixed, the gradient project algorithm (GPA; Jennrich, 2001, 2002) is applied for each group  $i = 1, \dots, I$  to minimize  $R_i^{SS}$  by imposing diagonality on Equation 2.
3. *Reflection and permutation*: The factors of group 1 are ordered according to their explained variance and reflected such that (most) strong loadings have a positive sign. Then, the factors of group  $i = 2, \dots, I$  are permuted and reflected to minimize the applied agreement criterion with the factor loadings of group 1 (i.e.,  $R_{i'}^A$  with  $i' = 1$ ).
4. *Constrained ML estimation*: The factor loadings and (co)variances are updated by maximizing the objective function  $LL + \lambda \times \text{vec}(F_{mg})$ , where  $\lambda$  is a vector of Lagrange multipliers and  $F_{mg}$  contains all group-specific restrictions  $F_{mg,i}$  (Equation 6) and is transformed into a vector by the ‘vec’ operator. Fisher scoring (Lee & Jennrich, 1979) is used, with possible step size

adjustments to prevent inadmissible factor covariance matrices, until the updates converge to a solution with both  $\hat{\lambda}$  and  $F_{mg}$  equal to zero, i.e., the (optimally rotated) ML solution.

### 3. Simulation study

#### 3.1. Problem

The goal of the simulation study is to evaluate the performance of MGFR with respect to the recovery of the (agreement and simple structure of the) factor loadings by the optimal rotation and the false positives (FP) and false negatives (FN) of hypothesis testing – based on the optimal rotation – for loading differences and non-zero loadings. For the rotation, we use the GP and LA criteria (Section 2.3.2.1) as  $R^A$  and oblimin (Equation 3) as  $R_i^{SS}$  for  $i = 1, \dots, I$ , with a variety of weights  $w$ . For the hypothesis testing, we focus on Wald tests because they are part of the default output of LG. We manipulated six factors that were expected to affect MGFR and/or the hypothesis testing: (1) the number of groups, (2) the group sizes, (3) the number of factors, (4) the type and size of the loading differences, and (5) the number of loading differences.

In terms of their effect on the performance, we hypothesize the following: (1) pertains to the number of loading matrices that are rotated and thus contributes to the complexity of the rotation. Together with (2), it defines the total sample size and thus the power for the Wald tests. The sampling fluctuations of group-specific factor loadings and factor covariance matrices are also determined by (2), which affect both the recovery and Wald tests. The number of factors (3) affects the complexity of the rotation and the power of the Wald tests, where a higher number of factors – given the same number of variables – implies a lower factor overdetermination and power (Hogarty et al., 2005). The difficulty of the rotation is further affected by (4) and (5) through the degree of the simple structure violations and disagreement between the groups. The power of the Wald tests for loading differences is also affected by the size of the differences.

### 3.2. Design and procedure

These factors were systematically varied in a complete factorial design:

1. the *number of groups* at 3 levels: 2, 4, 6;
2. the *group sizes* (i.e., number of observations per group) at 3 levels: 200, 600, 1000;
3. the *number of factors* at 2 levels: 2, 4;
4. the *type and size of loading differences* at 5 levels: primary loading shift, crossloading of .40, crossloading of .20, primary loading decrease of .40, primary loading decrease of .20;
5. the *number of loading differences* at 2 levels: 4, 16;

The group-specific factor loadings are all based on the same simple structure. In this ‘base matrix’, the fixed number of variables (i.e., 20) are equally distributed over the factors, i.e., each factor gets 10 non-zero loadings when  $Q = 2$  (Table 1) and five non-zero loadings when  $Q = 4$  (Table 2). Given that the unique variances vary around .40 (see below), the non-zero loadings are equal to  $\sqrt{.60}$  to obtain total variances of around one. From the common base, two different group-specific loading matrices are derived, each of which will pertain to half of the groups. Specifically, depending on the type and number of loading differences, for each of these two loading matrices, loadings were altered for a different set of variables (Tables 1, 2), referred to as ‘DIF items’. In case of a primary loading shift (PLS), two differences are induced per DIF item and thus one DIF item is selected per loading matrix to obtain a total of four loading differences across groups, or four DIF items (equally distributed across factors) are selected per loading matrix to obtain a total of 16 loading differences<sup>3</sup>. In particular, when  $Q = 2$ , the loadings  $[\sqrt{.6} \ 0]$  of the base matrix would be replaced by  $[0 \ \sqrt{.6}]$  (Table 1). When  $Q = 4$ , primary loadings are shifted similarly between factors 1 and 2 on the one hand, and between factors 3 and 4 on the other hand (Table 2);

---

<sup>3</sup> Note that when inducing >16 loading differences, the differences could be partially cancelled out by permuting factors (in case of PLS), increasing factor correlations (in case of CL) or rescaling factors (in case of PLD).

e.g.,  $[\sqrt{.6} \ 0 \ 0 \ 0]$  would become  $[0 \ \sqrt{.6} \ 0 \ 0]$ . For the crossloading differences (CL) and primary loading decreases (PLD), one loading was altered per DIF item and thus two DIF items are selected per loading matrix to obtain four differences across groups, or 8 to obtain 16 differences. In case of CL, the loadings  $[\sqrt{.6} \ 0 \ (0) \ (0)]$  would become  $[\sqrt{.6} \ .4 \ (0) \ (0)]$  or  $[\sqrt{.6} \ .2 \ (0) \ (0)]$  depending on the size of the CL. Note that a CL of .20 may be considered ‘ignorable’, whereas a CL of .40 is not (Stevens, 1992). To manipulate a PLD, the loadings  $[\sqrt{.6} \ 0 \ (0) \ (0)]$  would become  $[\sqrt{.6}-.4 \ 0 \ (0) \ (0)]$  or  $[\sqrt{.6}-.2 \ 0 \ (0) \ (0)]$  depending on the size of the decrease. Note that a PLD of .40 is considered a large non-invariance (Stark, Chernyshenko, & Drasgow, 2006) that can lead to incorrect statistical inference and biased parameter estimates (Hancock, Lawrence, & Nevitt, 2000). When  $I > 2$ , each of the two generated loading matrices was assigned to a random half of the groups. A number of remarks are in order: Firstly, in the case of four loading differences, only factors 1 and 2 are affected, even when  $Q = 4$ . Secondly, a PLS maintains the item’s communality whereas a CL increases it and a PLD lowers it. Thirdly, and most importantly, PLS and CL are difference types that cannot be readily traced by CFA-based methods such as multigroup BSEM or multigroup factor alignment (i.e., violations of configural invariance).

[ Insert Tables 1 and 2 about here]

The group-specific factor correlations are randomly sampled from a uniform distribution between  $-.50$  and  $.50$ , i.e.,  $U(-.50,.50)$ , and factor variances from  $U(.50,1.50)$ . Whenever a resulting  $\Psi_i$  is not positive definite, the sampling is repeated. Group-specific unique variances (i.e., diagonal of  $\mathbf{D}_i$ ) are sampled from  $U(.20,.60)$ . Factor scores are sampled from  $MVN(\mathbf{0}, \Psi_i)$  and residuals from  $MVN(\mathbf{0}, \mathbf{D}_i)$ , according to the specified group sizes. The group size of 200 corresponds to the recommended minimal sample size for obtaining accurate factor loading



estimates when item communalities are moderate (Fabrigar, MacCallum, Wegener, & Strahan, 1999; MacCallum, Widaman, Zhang, & Hong, 1999), whereas 1000 delimits a range of group sizes that largely corresponds to previous MI studies (Asparouhov & Muthén, 2014; Meade & Lautenschlager, 2004). Finally, the simulated data are created according to Equation 1. Note that the intercepts  $\tau_i$  are zero, since the focus is on loading differences.

According to this procedure, 50 data sets were generated per cell of the design, using Matlab R2017a. Thus, 3 (number of groups)  $\times$  3 (group sizes)  $\times$  2 (number of factors)  $\times$  5 (type/size of loading differences)  $\times$  2 (number of loading differences)  $\times$  50 (replications) = 9 000 data sets were generated. The data were analyzed by LG 6.0, using syntaxes (Appendix A). Since MGFR was applied with several  $R_{mg}$  criteria, one set of unrotated ML estimates was obtained and used as starting values for the optimal rotation (Steps 2 – 4; Section 2.2.3) per criterion. The average CPU time for multigroup EFA without rotation was 12s on an i7 processor with 8GB RAM. For 3 data sets, this estimation was repeated because it failed to converge the first time. Then, the following rotation criteria were applied – where ‘ $O$ ’ refers to oblimin:  $.01GP + .99O$ ,  $.10GP + .90O$ ,  $.30GP + .70O$ ,  $.50GP + .50O$ ,  $.70GP + .30O$ ,  $.01LA + .99O$ ,  $.10LA + .90O$ ,  $.30LA + .70O$ ,  $.50LA + .50O$ ,  $.70LA + .30O$ . The average CPU time of the rotation was 12s per criterion.

### **3.3. Results**

In this section, we first discuss the convergence of the optimal rotation per criterion (Section 3.3.1). Next, the recovery of the rotated loadings (Section 3.3.2) and corresponding factor (co)variances (Section 3.3.3) is discussed. Then, we present Wald test results based on the rotated loadings – for significant loading differences (Section 3.3.4) and loadings (Section 3.3.5). We end with conclusions and recommendations for empirical practice (Section 3.4).

#### *3.3.1. Convergence of optimal rotation identification*

Given that the optimal rotation is pursued by Fisher scoring, convergence is not guaranteed when initial values are far from the optimal rotation. Therefore, we first evaluate how often the procedure converges for each rotation criterion. Initially, the percentage of data sets for which the rotation converged,  $\%_{conv}$ , was 92.4%, 96.6%, 96.1%, 91.9%, and 82.4% when  $R^A = GP$  and  $w = .01, .10, .30, .50$  and  $.70$ , respectively. When  $R^A = LA$ , the  $\%_{conv}$ -values were 90.9%, 76.5%, 61.5%, 49.6%, and 40.4%, respectively. Since LA only has an acceptable  $\%_{conv}$  with  $w = .01$  – which did not improve by increasing  $\varepsilon$  – the other LA-based criteria are omitted from the results below. LA is based on square roots rather than squares of loading differences and, thus, has a larger impact on  $R_{mg}$  than GP. Therefore, it requires a smaller  $w$  to properly identify the group-specific factor (co)variances while maintaining simple structure per group. After re-running the non-converged rotations once, starting from a different random rotation of the loadings, the total  $\%_{conv}$  increased 2 to 5%. In Table 3, these  $\%_{conv}$  are given for the six rotations with the highest overall  $\%_{conv}$ , in function of the simulated conditions. Clearly,  $\%_{conv}$  is affected most by  $Q$ , with  $\%_{conv}$  equal to or near 100% when  $Q = 2$ . The ‘.70GP + .30O’ rotation has a markedly lower  $\%_{conv}$  for  $Q = 4$  than the other criteria. Thus, for comparability reasons, this criterion is also omitted from the results discussed below. The following results are based on the converged rotations only.

[ Insert Table 3 about here ]

### 3.3.2. Goodness-of-recovery of optimally rotated loadings

The recovery of the optimally rotated loading matrices is quantified by a goodness-of-loading-recovery statistic (*GOLR*), i.e., by computing congruence coefficients  $\varphi$  (Tucker, 1951) between the true ( $\lambda_{iq}$ ) and estimated ( $\hat{\lambda}_{iq}$ ) loadings and averaging across factors  $q$  and groups  $i$ :

$$GOLR = \frac{\sum_{i=1}^I \sum_{q=1}^Q \varphi(\lambda_{iq}, \hat{\lambda}_{iq})}{IQ}. \quad (10)$$

The *GOLR* evaluates the proportional equivalence of loadings (i.e., insensitive to factor rescaling) and varies between 0 (no agreement) and 1 (perfect agreement). Per criterion, the average *GOLR* is .99 ( $SD = .01$ ). This excellent recovery is hardly affected by the conditions.

Additionally, we inspect some more absolute – rather than proportional – measures to evaluate the recovery of the rotated loadings. The recovery of the simple structure is quantified by the mean estimated size of loadings that are zero in the population. On average, these mean estimates are quite small (about .05) and differences between criteria and conditions are subtle, except for an increase in case of PLS when GP is used with a higher weight; i.e., .03 for  $w = .01$ , .05 for  $w = .10$ , .08 for  $w = .30$ , and .12 for  $w = .50$ . To evaluate the recovery of between-group agreement, we calculated, per pair of groups, the mean difference between the estimated values of loadings that are invariant in the population and averaged this difference across group pairs. On average, it amounts to about .05 and it hardly varies across criteria or conditions.

### 3.3.3. Goodness-of-recovery of factor variances and covariances

To quantify the recovery of the factor (co)variances, the mean absolute difference (*MAD*) between the true ( $\psi_{iqq'}$ ) and estimated ( $\hat{\psi}_{iqq'}$ ) factor (co)variances is calculated as follows:

$$MAD_{\Psi} = \frac{\sum_{i=1}^I \sum_{q=1}^Q \sum_{q'=q}^Q |\psi_{iqq'} - \hat{\psi}_{iqq'}|}{IQ(Q+1)/2}. \quad (11)$$

The average  $MAD_{\Psi}$  -values in function of the criteria and conditions are given in Table 4. They vary around .07 or .08, indicating an overall good recovery of the  $\Psi_i$  matrices by each criterion.

[ Insert Table 4 about here ]

### 3.3.4. Wald tests for significant factor loading differences

To be conservative, we use .01 as the significance level<sup>4</sup>  $\alpha$  and the Bonferroni correction for multiple testing (Bonferroni, 1936), i.e., we divided  $\alpha$  by  $J \times Q$  and consider a loading to differ significantly when, for the corresponding Wald test,  $p < .00025$  for  $Q = 2$  and  $p < .000125$  for  $Q = 4$ . Table 5 presents percentages of data sets for which the Wald tests were perfectly correct (% correct; i.e., no FP and no FN), without FP (0 FP) and without FN (0 FN). For the % correct, we conclude that: (1) Overall, the ‘.50GP + .50O’ rotation gives the best results. (2) As an exception, for PLS, ‘.01LA + .99O’ performs better. (3) For PLD .40 and .20, ‘.10GP + .90O’ and ‘.30GP + .70O’ perform very similar to ‘.50GP + .50O’. (4) The lowest % correct are, not surprisingly, observed for the smallest CL and PLD of .20. (5) The performance is better in case of more groups, more observations per group, less factors and less differences (Section 3.1).

[ Insert Table 5 about here ]

When inspecting the ‘0 FP’ and ‘0 FN’ percentages, it is clear that: (1) For CL and PLD of .20, the lower % correct is mainly due to FN. (2) With an increasing  $I$  and  $N_i$ , we observe the well-known trade-off between FP and FN in function of sample size. (3) In case of more factors and more loading differences, the ‘0 FN’ and ‘0 FP’ percentages both decrease, indicating that the rotation is more intricate in these cases. Specifically, more factor variances and covariances need to be optimized when  $Q = 4$  and 16 differences make it challenging to pursue agreement and/or simple structure per group – the latter is true for 16 CL differences in particular. (4) Focusing on the best criterion per type/size of loading differences, the occurrence of FP is notably higher for CL .40. This confirms the suboptimal performance of oblimin – and most simple structure criteria – in case of item complexities larger than one (Lorenzo-Seva, 1999).

---

<sup>4</sup> The results for a (Bonferroni-corrected) significance level of .05 may be requested from the first author.

In Section 2.2.2.1, we pointed out that using GP as  $R^A$  could result in (false) small differences in an attempt to minimize (true) large differences, whereas LA eliminates (true) small loading differences while tolerating large differences. This explains why ‘.01LA + .99O’ performs best in case of PLS differences, which are the largest differences (i.e.,  $\sqrt{6} = .77$ ), and why ‘.50GP + .50O’ performs better for the other differences (i.e., of size .40 or .20). This is supported by the fact that ‘.50GP + .50O’ often results in FP for PLS differences (see Table 5).

Focusing on the best performing rotation for each type/size of loading differences (specified above), we inspected how many FP/FN occurred for each affected data set. Out of the 614 data sets with FP, only one FP was found for 401 (65%) and two FP for 97 data sets (16%). Out of the 1799 data sets with FN, only one FN was found for 465 (26%) and two FN for 308 data sets (17%). Of course, the FN are mainly found for the differences of .20.

To evaluate how MGFR performs in case of no differences, we performed an additional simulation study according to the procedure described above, without manipulating loading differences (i.e., retaining manipulations 1-3). Out of these 900 data sets, 97% of the converged ‘.50GP + .50O’ rotations resulted in zero FP, whereas for 89 data sets this rotation did not converge. Note that ‘.01LA + .99O’ failed to converge for 99% of these data sets.

### 3.3.5. Wald tests for significant factor loadings

For evaluating the MM(s) of the groups, we look at Wald tests for significance of factor loadings across groups<sup>5</sup>. Again, we focus on  $\alpha = .01$  and the Bonferroni correction divides  $\alpha$  by  $J \times Q$ . The percentage of data sets without FN does not differ across rotation criteria and is most

---

<sup>5</sup> The output of LG 6.0 also contains z tests per group. In this case, the Bonferroni correction divides  $\alpha$  by  $J \times Q \times I$ , which implies a loss in power. For these tests, the results on FP are highly similar as described in Section 3.3.5. The percentage of data sets without FN is lower, however, and this is especially the case for CL .40, CL .20 and PLD .40. Specifically, with ‘.50GP + .50O’ rotation, it is 80% for CL .40, 26% for CL .20, 80% for PLD .40. In practice, the results of the Wald tests for non-zero loadings *across* groups can be used to selectively test for non-zero loadings per group (i.e., to determine for *which* groups they apply), thus warranting a less rigorous correction for multiple testing.

affected by the type of loading differences. For ‘.50GP + .50O’ and ‘.01LA + .99O’, no FN occurred for 99 to 100% of the data sets for PLS, PLD .40 and PLD .20. For CL .40, 92 to 93% of the data sets are without FN and, for CL .20, this is 60 to 61%. The results for the FP are more intricate and are detailed in Table 6. The most important conclusion is that both the percentage of data sets without FP and the best performing rotation in this respect depend strongly on the type of loading differences. In case of PLS, GP with a higher weight appears to create more small crossloadings that are detected as FP, whereas ‘.01LA + .99O’ – also the preferred criterion for detecting PLS differences – performs very well with 96% of the data sets being free from FP. In case of CL .40 and CL .20, the best criterion for detecting the differences – i.e., ‘.50GP + .50O’ – is also the best one for avoiding FP non-zero loadings. The percentage of data sets without FP is still quite low – i.e., 42% and 56% for the CL of .40 and .20, respectively – again confirming that achieving simple structure is challenged by the CL. In case of PLD .40 and PLD .20, ‘.50GP + .50O’ is clearly suboptimal for detecting significant non-zero loadings whereas it is the best one for detecting the differences. Luckily, in Section 3.3.4., we found that ‘.10GP + .90O’ performed nearly the same in terms of revealing differences while it is the best one to avoid FP loadings in case of PLD. Selecting the mentioned best criterion for each type of loading differences, out of the 2085 data sets with FP, one FP was found for 751 (36%) and two for 384 (18%) of these data sets.

[ Insert Table 6 about here ]

### ***3.4. Conclusions and recommendations for empirical practice***

MGFR showed a good performance, especially given that the simulation study included small loading differences of .20. By means of the best rotation criterion for each configuration of loading differences, the loadings were recovered and rotated very well. Wald tests for detecting the differences were flawless (i.e., no FP and no FN) for roughly 70% of the data sets. When FP

or FN occurred, they often pertained to just one or two loadings. The simulation confirmed how the number of groups and group sizes make out the FN-FP trade-off. Furthermore, the performance drops somewhat in case of more factors and more differences, which make the rotation more challenging. It proved to be possible to evaluate the MM(s) at the same time, but, in case of CL, one should be aware of FP and, in case of PLD, a lower weight for GP is advised.

Since the best rotation criterion for detecting loading differences, as well as non-zero loadings, depends on the type and size of loading differences for a given data set, the following recommendations are in order: Because the type and size of loading differences are unknown beforehand and empirical data often contain a mix of differences, it is wise to first use the overall best criterion for distinguishing factor loading non-invariances; i.e., ' $.50GP + .50O$ '. Interestingly, this is equivalent to an unweighted combination of the agreement and simple structure criterion. Then, one could scrutinize the between-group differences of the obtained loadings and adjust the criterion as follows: (1) When the rotated solution reveals a few larger differences and many very small differences, it is advisable to see whether ' $.01LA + .99O$ ' eliminates the small ones. (2) When differences pertain to (rather subtle) PLD and one also wants to identify non-zero loadings, lowering  $w$  to .10 improves results for the latter while hardly affecting the detection of differences. (3) When differences pertain to CL, using LA or lowering  $w$  is not advisable. In this case, one may try whether an informed semi-specified target rotation (see Appendix A) improves the simple structure. (4) When a mix of differences occurs, the optimal choice is less clear-cut. Then, the advice is to resort to ' $.50GP + .50O$ ', but comparing to other criteria may still be informative.

#### **4. Application**

To illustrate the empirical value of MGFR, we applied it to data on the Open Sex Role Inventory (OSRI) downloaded from <https://openpsychometrics.org/rawdata/>. The OSRI is a

modernized measure of masculinity and femininity based on the Bem Sex Role Inventory (BSRI; Bem, 1974). Bem postulated that masculinity and femininity are two separate dimensions, allowing to characterize someone as masculine, feminine, androgynous or undifferentiated. The assumed MM of the BSRI has been widely contested, however (Choi & Fuqua, 2003). The OSRI contains 22 items (supposedly) measuring masculine characteristics alternated by 22 items measuring femininity (Appendix C). To the best of our knowledge, no studies on the MM of the OSRI have been published. Therefore, an EFA-based approach is preferred over CFA.

Note that the data is collected through the website and is thus not a random sample. For the purpose of our illustration, this is not a problem. Information is available on education, race, religion, gender and sexual orientation, as well as the country respondents are located in and whether English is their native language. We excluded non-native English speaking respondents to avoid differences due to misunderstanding items. Mainly respondents in the USA (2240), Great-Britain (357), Canada (180) and Australia (118) were left. Multigroup EFA confirmed factor loading invariance across gender, but revealed differences across sexual orientations and these results are reported below. Respondents with missing data on the items or grouping variable were excluded. For the reported analyses, 2767 respondents were included: 1539 hetero-, 568 bi-, 230 homo-, and 172 asexuals, and 258 who specified their sexuality as ‘other’.

The inadequacy of the masculine-feminine MM is confirmed by the fit of the corresponding baseline multigroup CFA model: CFI = .82 and RMSEA = .064. The CFI of multigroup EFA with two factors is .90 (RMSEA = .049) and dropped to .87 (RMSEA = .054) when imposing loading non-invariance. To identify the loading differences, MGFR was first applied with criterion ‘.50GP + .50O’ as recommended in Section 3.4. A mix of differences is found, corresponding to crossloadings appearing and primary loadings increasing or decreasing in one or more groups, but



differences are never as sizeable as the PLS in the simulation study (i.e., LA is not recommended). ‘.50GP + .50O’ rotation resulted in 14 loading differences and 71 non-zero loadings (out of 88), whereas ‘.10GP + .90O’ rotation resulted in 16 differences and 68 non-zero loadings, even though the rotated loadings look very similar. ‘.30GP + .70O’ rotation seemed to be a good middle ground with 14 differences and 69 non-zero loadings and these rotated loadings are given in Table 7, with Wald test  $p$ -values. Using simplimax-based group-specific targets did not improve the rotation.

Even though the factors can more or less be labelled ‘M’ (masculinity) and ‘F’ (femininity), hardly any of the items are pure measures of either M or F, which is supported by the  $p$ -values for the loadings (Table 7). Most of the significant loading differences seem to exist between heterosexuals on the one hand and (some of) the other groups on the other hand. This is confirmed by pairwise Wald tests that are obtained by the ‘knownclass’ option in LG (i.e., clustering the groups into five latent classes and enforcing a perfect prediction of class by group; Vermunt & Magidson, 2013). For example, for heterosexuals, Q4 (‘I give people handmade gifts’) has a negative CL on M and a lower primary loading for F. The factor covariance is non-significant for all groups:  $-.05$  for heterosexuals,  $.05$  for bisexuals,  $-.03$  for homosexuals,  $-.08$  for asexuals and  $-.04$  for ‘other’. The factor variances differ quite a bit across groups: the variances of M are 1.33, .98, .90, .89, .90, and the variances of F are .99, .89, 1.00, .88, 1.25 for that same order of groups, respectively. Therefore, oblimin rotation per group with fixed factor variances, using the Jennrich (1973) restrictions, overestimates the loading differences, i.e., 26 differences are found to be significant. In any case, before using the OSRI for comparing masculinity and femininity across sexual orientations, it needs to be revised to a large extent.

## 5. Discussion

Testing for MI is essential before comparing latent constructs across groups. When factor loading invariance fails, further MI tests are ruled out and one can either ignore the non-invariance and risk invalid conclusions, refrain from further analyses, or take action by scrutinizing loading differences. The latter may give clues on how non-invariances can be avoided in future research (e.g., excluding or rephrasing items). When looking for all kinds of differences (i.e., including primary loading shifts and crossloadings), multigroup EFA is the way to go. To properly identify these non-invariances, MGFR pursues both agreement and simple structure, disentangles loading differences from differences in the structural model, and enables hypothesis tests for the loadings.

When using the loading alignment criterion for agreement, MGFR may be conceived as an EFA extension of multigroup factor alignment (MGFA; Asparouhov & Muthén, 2014) in that it both aligns and rotates, albeit that – for now – it focuses on factor loadings only. Of course, before making latent construct comparisons, intercept invariance should be addressed as well, but like in MI testing, we prefer to tackle the levels of MI in a stepwise manner. While MGFA only assesses primary loadings and assumes differences to be small and pertaining to a minority of the loadings or groups (i.e., partial and/or approximate MI), we are not even assuming an invariant zero loading pattern. Therefore, it makes no sense to align the intercepts for enabling factor mean comparisons *while* rotating the factors toward one another to assess whether they are somewhat comparable in the first place. In future research, it would be interesting to study how MGFR can be combined with intercept alignment and whether it indeed needs to be a stepwise approach. To this end, the principles of MGFA need to be extended to the multi-factor EFA case, whereas currently it cannot even align CFA models with a crossloading. Clearly, the latter warrants a separate study in itself.

Since MGFR proved to be very promising, it would be worthwhile to devote more research to refining and extending it in a number of respects. Firstly, it would be interesting to determine

invariant sets (Asparouhov & Muthén, 2014) of groups per factor loading, building on the pairwise Wald tests mentioned in Section 4. Secondly, the unrotated solution that is fed to the rotation procedure (Section 2.2.3) corresponds to a single set of random ‘starting values’ for the rotation and the latter may end up in a local optimum depending on these values. Future research will include an evaluation of the sensitivity to local optima and, if necessary, the development of a multistart MGFR procedure or a multigroup extension of the GPA, compatible with free factor variances. For now, the user is advised to repeat the analysis a few times to see whether this affects results. Thirdly, the rotation depends on the weight of the agreement versus the simple structure criterion. The best weight to use depends on the loading differences. It would be interesting to evaluate whether it can be automatically optimized for comparing the loadings for a given data set. Until then, the user is advised to compare a few rotations (Section 3.4).

Finally, an interesting question is to what extent MGFR can serve as a precursor to multigroup EFA or CFA with partial loading invariance according to the identified loading differences and MM(s). Needless to say, this requires a crossvalidation approach (Gerbing & Hamilton, 1996), e.g., where each group is split in random halves, and thus larger sample sizes. When group sizes are too small or the number of groups is large, MGFR can team up with a mixture approach such as proposed by De Roover et al. (2017), where groups are clustered according to the similarity of their loadings and the rotation would be applied per cluster.

### References

- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397-438.
- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 495-508.
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*, 155.
- Bonferroni, C. E. (1936). *Teoria statistica delle classi e calcolo delle probabilita*. Libreria internazionale Seeber.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate behavioral research*, *36*, 111-150.
- Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory factor model. *Structural Equation Modeling*, *16*, 295-314.
- Choi, N., & Fuqua, D. R. (2003). The structure of the Bem Sex Role Inventory: A summary report of 23 validation studies. *Educational and Psychological Measurement*, *63*, 872-887.
- Clarkson, D. B., & Jennrich, R. I. (1988). Quartic rotation criteria and algorithms. *Psychometrika*, *53*, 251-259.
- 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.
- Ferrando, P. J., & Lorenzo Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicológica*, *21*, 301-323.

- Gerbing, D. W., & Hamilton, J. G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 3, 62-72.
- Hambleton, R. K. (2006). Good practices for identifying differential item functioning. *Medical Care*, 44, S182-S188.
- Hamilton, L. C. (2012). *Statistics with Stata: version 12*. Cengage Learning.
- Hancock, G. R., Lawrence, F. R., & Nevitt, J. (2000). Type I error and power of latent mean methods and MANOVA in factorially invariant and noninvariant latent variable systems. *Structural Equation Modeling*, 7, 534-556.
- Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Mathematical and Statistical Psychology*, 17, 65-70.
- Hessen, D. J., Dolan, C. V., & Wicherts, J. M. (2006). Multi-group exploratory factor analysis and the power to detect uniform bias. *Applied Psychological Research*, 30, 233-246.
- Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M. & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, 65, 202-226.
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Psychology Press.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jennrich, R. I. (1973). Standard errors for obliquely rotated factor loadings. *Psychometrika*, 38, 593-604.

- Jennrich, R. I. (1979). Admissible values of  $\gamma$  in direct oblimin rotation. *Psychometrika*, *44*, 173-177.
- Jennrich, R. I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika*, *66*, 289-306.
- Jennrich, R. I. (2002). A simple general method for oblique rotation. *Psychometrika*, *67*, 7-19.
- Jennrich, R. I. & Sampson, P. F. (1966). Rotation to simple loadings.. *Psychometrika*, *31*, 313-323.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409-426.
- Jöreskog, K. G. (1979). A general approach to confirmatory maximum likelihood factor analysis, with addendum. In K. G. Jöreskog & D. Sörbom, *Advances in factor analysis and structural equation models* (pp. 21-43). Cambridge, MA: Abt Books.
- Lawley, D. N., & Maxwell, A. E. (1962). Factor analysis as a statistical method. *The Statistician*, *12*, 209-229.
- Kiers, H. A. (1994). Simplimax: Oblique rotation to an optimal target with simple structure. *Psychometrika*, *59*, 567-579.
- Lee, S. Y., & Jennrich, R. I. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika*, *44*, 99-113.
- Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research*, *34*, 347-365.
- Lorenzo-Seva, U. (2000). The weighted oblimin rotation. *Psychometrika*, *65*, 301-318.
- Lorenzo-Seva, U., Kiers, H. A., & Berge, J. M. (2002). Techniques for oblique factor rotation of two or more loading matrices to a mixture of simple structure and optimal agreement. *British Journal of Mathematical and Statistical Psychology*, *55*, 337-360.

- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological bulletin*, *111*, 490.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological methods*, *4*, 84.
- Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual review of clinical psychology*, *10*, 85-110.
- McCrae, R. R., Zonderman, A. B., Costa Jr, P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*, 552.
- Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling*, *11*, 60-72.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543.
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological methods*, *17*, 313.
- Muthén, B., & Asparouhov, T. (2013). BSEM measurement invariance analysis. *Mplus Web Notes*, *17*, 1-48.
- Muthén, L. K., & Muthén, B. O. (2005). *Mplus: Statistical analysis with latent variables: User's guide*. Los Angeles: Muthén & Muthén.

- Neale, M. C., Boker, S. M., Xie, G. & Maes, H. H. (2003). *Mx: Statistical modeling, 6th ed.* Richmond, VA: Virginia Commonwealth University, Department of Psychiatry.
- Nie, N. H., Bent, D. H., & Hull, C. H. (1970). *SPSS: Statistical package for the social sciences* (No. HA29 S6). New York: McGraw-Hill.
- Osborne, J. W. (2015). What is rotating in exploratory factor analysis. *Practical Assessment, Research & Evaluation, 20*, 1-7.
- Pennell, R. (1968). The influence of communality and N on the sampling distributions of factor loadings. *Psychometrika, 33*, 423-439.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software, 48*, 1-36.
- Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of cross-cultural psychology, 38*, 173-212.
- Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement, 71*, 95-113.
- Sireci, S. G., & Rios, J. A. (2013). Decisions that make a difference in detecting differential item functioning. *Educational Research and Evaluation, 19*, 170-187.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229–239.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*, 1292-1306.



- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- ten Berge, J. M. (1977). Orthogonal Procrustes rotation for two or more matrices. *Psychometrika*, *42*, 267-276.
- Teresi, J. A., & Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research*, *16*, 33-42.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: The University of Chicago Press.
- Tucker, L. R. (1951). A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984). Washington, DC: Department of the Army.
- Vermunt, J. K., & Magidson, J. (2013). *Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations Inc.
- Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. Albany, NY: State University of New York Press.

Table 1. Base loading matrix and the derived group-specific loading matrices, in case of two factors. For each of the group-specific loading matrices, the affected loadings for each type of loading differences are indicated as follows: ‘PLS’ refers to primary loading shift, ‘CL’ to crossloading and ‘PLD’ to primary loading decrease. Differences between brackets are only induced in the case of 16 loading differences.

	Base loading matrix		Group-specific loading matrix 1		Group-specific loading matrix 2	
	F1	F2	F1	F2	F1	F2
V1	$\sqrt{.6}$	0	<i>PLS, PLD</i>	<i>PLS, CL</i>	$\sqrt{.6}$	0
V2	$\sqrt{.6}$	0	<i>(PLS), (PLD)</i>	<i>(PLS), (CL)</i>	$\sqrt{.6}$	0
V3	$\sqrt{.6}$	0	<i>(PLD)</i>	<i>(CL)</i>	<i>PLS</i>	<i>PLS</i>
V4	$\sqrt{.6}$	0	<i>(PLD)</i>	<i>(CL)</i>	<i>(PLS)</i>	<i>(PLS)</i>
V5	$\sqrt{.6}$	0	$\sqrt{.6}$	0	<i>PLD</i>	<i>CL</i>
V6	$\sqrt{.6}$	0	$\sqrt{.6}$	0	<i>(PLD)</i>	<i>(CL)</i>
V7	$\sqrt{.6}$	0	$\sqrt{.6}$	0	<i>(PLD)</i>	<i>(CL)</i>
V8	$\sqrt{.6}$	0	$\sqrt{.6}$	0	<i>(PLD)</i>	<i>(CL)</i>
V9	$\sqrt{.6}$	0	$\sqrt{.6}$	0	$\sqrt{.6}$	0
V10	$\sqrt{.6}$	0	$\sqrt{.6}$	0	$\sqrt{.6}$	0
V11	0	$\sqrt{.6}$	<i>(PLS), CL</i>	<i>(PLS), PLD</i>	0	$\sqrt{.6}$
V12	0	$\sqrt{.6}$	<i>(PLS), (CL)</i>	<i>(PLS), (PLD)</i>	0	$\sqrt{.6}$
V13	0	$\sqrt{.6}$	<i>(CL)</i>	<i>(PLD)</i>	<i>(PLS)</i>	<i>(PLS)</i>
V14	0	$\sqrt{.6}$	<i>(CL)</i>	<i>(PLD)</i>	<i>(PLS)</i>	<i>(PLS)</i>
V15	0	$\sqrt{.6}$	0	$\sqrt{.6}$	<i>CL</i>	<i>PLD</i>
V16	0	$\sqrt{.6}$	0	$\sqrt{.6}$	<i>(CL)</i>	<i>(PLD)</i>
V17	0	$\sqrt{.6}$	0	$\sqrt{.6}$	<i>(CL)</i>	<i>(PLD)</i>
V18	0	$\sqrt{.6}$	0	$\sqrt{.6}$	<i>(CL)</i>	<i>(PLD)</i>
V19	0	$\sqrt{.6}$	0	$\sqrt{.6}$	0	$\sqrt{.6}$
V20	0	$\sqrt{.6}$	0	$\sqrt{.6}$	0	$\sqrt{.6}$

Table 2. Base loading matrix and the derived group-specific loading matrices, in case of four factors. For each of the group-specific loading matrices, the affected loadings for each type of loading differences are indicated as follows: ‘PLS’ refers to primary loading shift, ‘CL’ to crossloading and ‘PLD’ to primary loading decrease. Differences between brackets are only induced in the case of 16 loading differences.

	Base loading matrix				Group-specific loading matrix 1				Group-specific loading matrix 2			
	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4
V1	$\sqrt{6}$	0	0	0	<i>PLS</i> , <i>PLD</i>	<i>PLS</i> , <i>CL</i>	0	0	$\sqrt{6}$	0	0	0
V2	$\sqrt{6}$	0	0	0	<i>(PLD)</i>	<i>(CL)</i>	0	0	<i>PLS</i>	<i>PLS</i>	0	0
V3	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	0	<i>PLD</i>	<i>CL</i>	0	0
V4	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	0	<i>(PLD)</i>	<i>(CL)</i>	0	0
V5	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	0
V6	0	$\sqrt{6}$	0	0	<i>(PLS)</i> , <i>CL</i>	<i>(PLS)</i> , <i>PLD</i>	0	0	0	$\sqrt{6}$	0	0
V7	0	$\sqrt{6}$	0	0	<i>(CL)</i>	<i>(PLD)</i>	0	0	<i>(PLS)</i>	<i>(PLS)</i>	0	0
V8	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	<i>CL</i>	<i>PLD</i>	0	0
V9	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	<i>(CL)</i>	<i>(PLD)</i>	0	0
V10	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0
V11	0	0	$\sqrt{6}$	0	0	0	<i>(PLS)</i> , <i>(PLD)</i>	<i>(PLS)</i> , <i>(CL)</i>	0	0	$\sqrt{6}$	0
V12	0	0	$\sqrt{6}$	0	0	0	<i>(PLD)</i>	<i>(CL)</i>	0	0	<i>(PLS)</i>	<i>(PLS)</i>
V13	0	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	0	<i>(PLD)</i>	<i>(CL)</i>
V14	0	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	0	<i>(PLD)</i>	<i>(CL)</i>
V15	0	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0
V16	0	0	0	$\sqrt{6}$	0	0	<i>(PLS)</i> , <i>(CL)</i>	<i>(PLS)</i> , <i>(PLD)</i>	0	0	0	$\sqrt{6}$
V17	0	0	0	$\sqrt{6}$	0	0	<i>(CL)</i>	<i>(PLD)</i>	0	0	<i>(PLS)</i>	<i>(PSL)</i>
V18	0	0	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	<i>(CL)</i>	<i>(PLD)</i>
V19	0	0	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	<i>(CL)</i>	<i>(PLD)</i>
V20	0	0	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$	0	0	0	$\sqrt{6}$

Table 3. Convergence frequencies (%) of the optimal rotation procedure for six criteria, in function of the simulated conditions. ‘GP’ = generalized procrustes, ‘LA’ = loading alignment and ‘O’ = oblimin.

	.01GP + .99O	.10GP + .90O	.30GP + .70O	.50GP + .50O	.70GP + .30O	.01LA + .99O
<i>I</i> =2	96.0%	98.4%	98.7%	98.4%	96.7%	96.8%
<i>I</i> =4	94.3%	97.2%	96.8%	94.6%	87.7%	92.7%
<i>I</i> =6	94.3%	96.1%	95.7%	90.5%	76.8%	87.3%
<i>N<sub>i</sub></i> =200	96.7%	97.4%	97.3%	96.1%	91.9%	96.4%
<i>N<sub>i</sub></i> =600	95.4%	97.4%	97.5%	94.9%	86.0%	92.7%
<i>N<sub>i</sub></i> =1000	92.6%	96.9%	96.5%	92.4%	83.4%	87.7%
<i>Q</i> =2	100%	100%	100%	100%	99.9%	94.7%
<i>Q</i> =4	89.8%	94.5%	94.2%	89.0%	74.2%	89.8%
PLS	95.1%	97.8%	97.5%	95.5%	83.4%	89.8%
CL .40	95.7%	96.8%	97.0%	93.4%	86.1%	94.9%
CL .20	95.9%	97.5%	96.9%	94.8%	89.1%	96.1%
PLD .40	94.1%	96.9%	97.1%	95.6%	90.1%	89.4%
PLD .20	93.8%	97.2%	96.9%	93.2%	86.7%	91.1%
4 diff.	94.4%	96.9%	96.9%	94.2%	84.9%	89.9%
16 diff.	95.4%	97.6%	97.2%	94.7%	89.3%	94.6%
Total	94.9%	97.2%	97.1%	94.5%	87.1%	92.3%

Table 4. Mean absolute difference between the true and estimated factor variances and covariances, in function of five rotation criteria and the simulated conditions.

	.01GP + .99O	.10GP + .90O	.30GP + .70O	.50GP + .50O	.01LA + .99O
<i>I</i> =2	.08	.06	.07	.07	.07
<i>I</i> =4	.08	.07	.07	.08	.07
<i>I</i> =6	.08	.07	.08	.09	.07
<i>N<sub>i</sub></i> =200	.10	.08	.09	.10	.09
<i>N<sub>i</sub></i> =600	.08	.06	.07	.07	.06
<i>N<sub>i</sub></i> =1000	.07	.06	.06	.07	.06
<i>Q</i> =2	.08	.07	.08	.08	.07
<i>Q</i> =4	.08	.07	.07	.08	.07
PLS	.06	.07	.11	.15	.05
CL .40	.13	.10	.08	.08	.11
CL .20	.09	.07	.07	.07	.08
PLD .40	.06	.05	.05	.06	.06
PLD .20	.06	.05	.05	.05	.06
4 diff.	.07	.05	.06	.07	.06
16 diff.	.09	.08	.09	.09	.08
Total	.08	.07	.07	.08	.07

Table 5. Percentages (%) of data sets for which the Wald test results ( $\alpha = .01$ , Bonferroni corrected) for between-group loading differences are perfectly correct (i.e., no false positives and no false negatives;  $\hat{b}$ ), without false positives (0 FP) and without false negatives (0 FN). For each simulated condition, the best  $\hat{b}$ % is indicated in bold face.

	.01GP + .990	.10GP + .900	.30GP + .700	.50GP + .500	.01LA + .990	.01GP + .990	.10GP + .900	.30GP + .700	.50GP + .500	.01LA + .990	0 FP	0 FN	0 FP	0 FN	0 FP	0 FN	0 FP	0 FN
<i>I</i> =2	43	56	60	<b>61</b>	49	80	57	89	65	92	66	93	67	90	56			
<i>I</i> =4	45	68	71	<b>74</b>	61	64	75	84	82	87	83	89	83	85	73			
<i>I</i> =6	44	67	71	<b>73</b>	64	56	82	77	88	82	88	84	88	81	80			
<i>N<sub>i</sub></i> =200	36	48	50	<b>50</b>	41	77	48	90	54	92	55	92	55	90	46			
<i>N<sub>i</sub></i> =600	48	70	74	<b>77</b>	64	65	79	83	86	87	86	89	87	85	77			
<i>N<sub>i</sub></i> =1000	48	73	78	<b>81</b>	70	58	87	78	95	82	95	86	96	81	87			
<i>Q</i> =2	56	72	76	<b>79</b>	69	73	81	88	84	92	84	94	84	90	78			
<i>Q</i> =4	30	55	58	<b>59</b>	46	60	60	79	72	82	74	82	73	80	60			
PLS	72	79	76	74	<b>93</b>	73	98	79	100	76	100	74	100	94	99			
CL .40	25	57	73	<b>81</b>	50	31	91	63	93	78	94	86	94	57	90			
CL .20	25	51	52	<b>54</b>	49	57	56	91	56	94	56	96	56	91	53			
PLD .40	67	84	86	<b>86</b>	70	86	76	92	91	93	92	92	93	93	74			
PLD .20	31	47	50	<b>50</b>	29	88	33	93	51	94	53	95	52	93	30			
4 diff.	53	74	76	<b>76</b>	68	72	75	90	82	93	82	92	82	90	76			
16 diff.	36	53	58	<b>62</b>	48	62	66	77	74	81	75	85	76	81	63			
Total	44	64	67	<b>69</b>	58	67	71	83	78	87	79	89	79	85	69			

Table 6. Percentages (%) of data sets for which the Wald test results ( $\alpha = .01$ , Bonferroni corrected) for significant loadings across groups are without false positives (0 FP).

	<i>.01GP + .99O</i>	<i>.10GP + .90O</i>	<i>.30GP + .70O</i>	<i>.50GP + .50O</i>	<i>.01LA + .99O</i>
<i>I=2</i>	76	71	64	62	76
<i>I=4</i>	68	56	50	49	69
<i>I=6</i>	63	47	41	40	63
<i>N<sub>i</sub>=200</i>	78	73	67	64	80
<i>N<sub>i</sub>=600</i>	67	54	48	48	67
<i>N<sub>i</sub>=1000</i>	63	47	40	39	61
<i>Q=2</i>	74	63	59	62	73
<i>Q=4</i>	64	54	45	38	67
PLS	95	38	09	03	96
CL .40	17	19	30	42	18
CL .20	48	50	52	56	50
PLD .40	94	92	84	73	94
PLD .20	93	92	85	79	95
4 diff.	78	70	61	57	78
16 diff.	61	46	43	44	61
Total	69	58	52	51	70

Table 7. ‘.30GP + .70O’ rotated loadings per sexual orientation for the OSRI data and  $p$ -values. ‘M’ refers to masculinity, ‘F’ to femininity, ‘Wald(=)’ to tests for loading differences and ‘Wald(0)’ to tests for non-zero loadings. P-values that are significant at a Bonferroni-corrected 1% significance level (i.e.,  $p < .00014$ ) are in bold face, as well as loadings that differ significantly across groups.

	Hetero- sexual		Bisexual		Homo- sexual		Asexual		Other		Wald(=) $p$ -values		Wald(0) $p$ -values	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Q1	.17	.03	.14	.10	-.06	.15	.07	.11	-.16	.18	.0007	.3100	<b>.0000</b>	.0020
Q2	<b>-.44</b>	.38	<b>-.23</b>	.54	<b>.00</b>	.56	<b>-.13</b>	.53	<b>.00</b>	.50	<b>.0000</b>	.1100	<b>.0000</b>	<b>.0000</b>
Q3	<b>.66</b>	-.19	<b>.76</b>	.00	<b>.52</b>	-.09	<b>.48</b>	-.05	<b>.37</b>	-.02	<b>.0000</b>	.0850	<b>.0000</b>	.0094
Q4	<b>-.42</b>	<b>.60</b>	<b>-.04</b>	<b>.94</b>	<b>-.09</b>	<b>1.09</b>	<b>-.08</b>	<b>1.06</b>	<b>-.04</b>	<b>1.06</b>	<b>.0000</b>	<b>.0000</b>	<b>.0000</b>	<b>.0000</b>
Q5	.47	-.01	.61	.22	.48	.17	.64	.15	.58	.03	.1000	.0130	<b>.0000</b>	<b>.0001</b>
Q6	-.14	.52	.02	.65	.05	.83	.12	.61	.03	.73	.0160	.0011	.0320	<b>.0000</b>
Q7	.21	.06	.07	.06	.11	.08	.12	-.03	.20	.15	.4100	.3200	<b>.0000</b>	.0620
Q8	-.27	.36	-.18	.50	-.16	.49	-.16	.47	-.14	.56	.4900	.1500	<b>.0000</b>	<b>.0000</b>
Q9	.43	-.21	.32	-.26	.61	-.24	.50	-.27	.44	-.34	.0150	.5400	<b>.0000</b>	<b>.0000</b>
Q10	-.34	.56	-.37	.77	-.13	.72	-.16	.81	-.18	.72	.0057	.0110	<b>.0000</b>	<b>.0000</b>
Q11	<b>.32</b>	-.01	<b>.11</b>	.05	<b>-.08</b>	.03	<b>.19</b>	.00	<b>.15</b>	.10	<b>.0000</b>	.6200	<b>.0000</b>	.5500
Q12	-.16	.19	-.22	.29	-.26	.35	-.29	.17	-.13	.37	.4300	.0810	<b>.0000</b>	<b>.0000</b>
Q13	.55	-.27	.61	-.20	.67	-.28	.72	-.25	.74	-.26	.0950	.9100	<b>.0000</b>	<b>.0000</b>
Q14	-.51	.11	-.31	.13	-.40	.17	-.42	.20	-.15	.17	.0019	.8800	<b>.0000</b>	<b>.0000</b>
Q15	.63	-.09	.79	.14	.80	.04	.78	.06	.72	.13	.0970	.0092	<b>.0000</b>	.0042
Q16	-.06	.69	.10	.83	.10	.87	.13	.90	.00	.87	.0420	.0360	.0270	<b>.0000</b>
Q17	.64	-.17	.70	-.02	.79	-.10	.74	-.06	.72	-.13	.3900	.4200	<b>.0000</b>	.0020
Q18	-.27	.59	-.12	.70	-.04	.81	-.03	.82	.04	.75	.0025	.0340	<b>.0000</b>	<b>.0000</b>
Q19	.27	.05	.32	.18	.17	.08	.23	.08	.25	.24	.5700	.1200	<b>.0000</b>	<b>.0001</b>
Q20	-.35	.44	-.33	.50	-.37	.45	-.39	.32	-.39	.44	.9600	.4400	<b>.0000</b>	<b>.0000</b>
Q21	.74	-.21	.58	-.20	.49	-.27	.49	-.18	.52	-.05	.0003	.0290	<b>.0000</b>	<b>.0000</b>
Q22	-.31	.28	-.24	.29	-.52	.38	-.42	.36	-.33	.51	.0065	.0210	<b>.0000</b>	<b>.0000</b>
Q23	.49	-.24	.47	-.05	.61	-.09	.42	-.07	.54	-.06	.1600	.0410	<b>.0000</b>	<b>.0000</b>
Q24	-.08	.34	.06	.29	.03	.30	.01	.23	-.08	.38	.2400	.3800	<b>.3000</b>	<b>.0000</b>
Q25	.40	-.22	.33	-.26	.14	-.16	.32	-.11	.36	-.05	.0270	.0970	<b>.0000</b>	<b>.0000</b>
Q26	-.29	.40	-.25	.51	-.12	.45	-.24	.33	-.02	.34	.0041	.1100	<b>.0000</b>	<b>.0000</b>
Q27	.45	-.02	.63	.06	.63	-.03	.72	.12	.64	.04	.0051	.3000	<b>.0000</b>	.2900
Q28	<b>.20</b>	<b>1.45</b>	<b>.09</b>	<b>.87</b>	<b>-.17</b>	<b>.76</b>	<b>.02</b>	<b>1.05</b>	<b>-.18</b>	<b>.86</b>	<b>.0000</b>	<b>.0000</b>	<b>.0000</b>	<b>.0000</b>
Q29	.41	-.05	.47	-.06	.38	.13	.54	.03	.57	.21	.1900	.0033	<b>.0000</b>	.0021
Q30	-.22	<b>.51</b>	-.11	<b>.98</b>	-.05	<b>.83</b>	-.11	<b>.80</b>	-.09	<b>.83</b>	.2700	<b>.0000</b>	.0001	<b>.0000</b>
Q31	.50	-.06	.58	-.05	.51	-.03	.57	-.06	.34	-.07	.1300	.9900	<b>.0000</b>	.4700
Q32	-.34	.35	-.27	.27	-.49	.41	-.30	.27	-.49	.54	.0330	.0065	<b>.0000</b>	<b>.0000</b>
Q33	.57	-.02	.62	-.11	.39	-.09	.34	-.07	.44	-.12	.0280	.8000	<b>.0000</b>	.2600
Q34	-.34	.38	-.27	.27	-.32	.29	-.44	.20	-.49	.33	.0750	.2000	<b>.0000</b>	<b>.0000</b>
Q35	.51	-.14	.58	-.13	.63	-.13	.47	-.11	.68	.00	.1500	.3500	<b>.0000</b>	.0055
Q36	-.38	<b>.52</b>	-.18	<b>.82</b>	-.10	<b>.83</b>	-.21	<b>.80</b>	-.08	<b>.89</b>	.0004	<b>.0000</b>	<b>.0000</b>	.0000
Q37	<b>.23</b>	-.11	<b>.46</b>	-.05	<b>.50</b>	.03	<b>.57</b>	-.02	<b>.64</b>	-.19	<b>.0000</b>	.0910	<b>.0000</b>	.0058
Q38	-.44	.31	-.48	.53	-.24	.52	-.25	.47	-.26	.43	.0086	.0360	<b>.0000</b>	.0000
Q39	.51	-.19	.68	.00	.69	.10	.72	-.10	.69	-.07	.0110	.0008	<b>.0000</b>	.0001
Q40	<b>-.36</b>	<b>.70</b>	<b>-.04</b>	<b>1.13</b>	<b>-.03</b>	<b>1.04</b>	<b>-.30</b>	<b>1.03</b>	<b>-.01</b>	<b>.99</b>	<b>.0000</b>	<b>.0000</b>	<b>.0000</b>	.0000
Q41	.55	-.15	.80	-.06	.79	-.06	.73	-.13	.80	.03	.0012	.1100	<b>.0000</b>	.0074
Q42	-.28	.34	-.22	.48	-.09	.54	-.06	.41	-.08	.60	.0200	.0087	<b>.0000</b>	<b>.0000</b>
Q43	.47	-.06	.49	.05	.61	.20	.66	-.03	.59	.05	.1100	.0270	<b>.0000</b>	.0330
Q44	<b>.19</b>	<b>1.51</b>	<b>.08</b>	<b>.92</b>	<b>-.20</b>	<b>.79</b>	<b>.00</b>	<b>1.10</b>	<b>-.16</b>	<b>.86</b>	<b>.0000</b>	<b>.0000</b>	<b>.0000</b>	<b>.0000</b>

## Appendix A

An example syntax for a twenty-item four-factor multigroup EFA with optimal rotation is:

```
options
  algorithm
    tolerance=1e-008 emtolerance=0.01 emiterations=2500 nriterations=500;
  startvalues
    seed=0 sets=5 tolerance=1e-005 iterations=100 PCA;
  missing excludeall;
  efa oblimin procrustes=50;
  output
    iterationdetail classification parameters=effect standarderrors rotation
writeparameters='results_parameters.csv' write='results.csv' writeloadings='results_loadings.txt';
variables
  dependent V1 continuous, V2 continuous, V3 continuous, V4 continuous, V5 continuous, V6
continuous, V7 continuous, V8 continuous, V9 continuous, V10 continuous, V11 continuous, V12
continuous, V13 continuous, V14 continuous, V15 continuous, V16 continuous, V17 continuous, V18
continuous, V19 continuous, V20 continuous;
  independent G nominal;
  latent
    F1 continuous,
    F2 continuous,
    F3 continuous,
    F4 continuous;
equations
// factor variances and covariances
F1 | G;
F2 | G;
F3 | G;
F4 | G;
F1 <-> F2 | G;
F1 <-> F3 | G;
F1 <-> F4 | G;
F2 <-> F3 | G;
F2 <-> F4 | G;
F3 <-> F4 | G;
// regression models for items
V1 - V20 <- 1 | G + F1 | G + F2 | G + F3 | G + F4 | G;
// unique variances
V1 - V20 | G;
```



The categorical variable ‘G’ indicates the group memberships of the observations and ‘V1’ to ‘V20’ refer to the twenty items – they are to be replaced by the variable labels in the data set at hand. Details about the technical settings can be found in the Latent Gold manual (Vermunt & Magidson, 2013). ‘PCA’ refers to randomized PCA-based starting values that are described in De Roover, Vermunt, Timmerman, and Ceulemans (2017). Note that both the factor variances and covariances are free to vary across groups and that the optimal rotation is requested by ‘efa oblimin procrustes=50’. In general, the latter has the following structure:

```
efa <simple structure criterion> <agreement criterion>=<w*100>
```

The simple structure criterion (see Section 2.3.1.1) can be ‘oblimin’, ‘geomin’ or ‘varimax’ – where the latter is orthogonal and should be used with factor covariances equal to zero (i.e., deleting the ‘Fx <-> Fx | G’ lines in the syntax). The agreement criterion (see Section 2.3.2.1) can be either ‘procrustes’ for GP or ‘alignment’ for LA. When one wants to use alignment with a user-specified value for  $\varepsilon$  (the default is  $1 \times 10^{-12}$ ), the command becomes, e.g., ‘efa oblimin alignment=1 epsilon=1e-6’.

As an alternative simple structure criterion, target rotation can be applied by using ‘target=’filename.txt’’, where the file should contain *I* group-specific targets or one target to be used for all groups. Note that ‘-99’ is used to indicate non-specified parts of the targets. For instance, two semi-specified group-specific targets for eight items and two factors would be communicated as follows:

```
‘0 -99
0 -99
0 -99
0 -99
-99 0
-99 0
-99 0
```

-99 -99

-99 -99

0 -99

0 -99

0 -99

0 -99

-99 0

-99 0

-99 0'

To start from user-specified parameter values and only perform the rotation (e.g., to try different rotation criteria without repeating the model estimation), the 'algorithm' and 'startvalues' options can be modified as follows:

algorithm

tolerance=1e-008 emtolerance=0.01 emiterations=0 nriterations=0;

startvalues

seed=0 sets=1 tolerance=1e-005 iterations=0;

The user-specified parameter values are communicated through a text file containing the parameter values in the internal order of the parameters (Vermunt & Magidson, 2003), which is specified at the end of the syntax as 'startingvalues.txt'.

### Appendix B

When the unrotated factors of group  $i$  are orthonormal, the true (i.e., population-level) optimally rotated factor loadings  $\Lambda_i$  and factor covariance matrix  $\Psi_i$  can be expressed as functions of the unrotated orthonormal true loadings  $\mathbf{A}_i$  as follows:

$$\begin{aligned}\Lambda_i &= \mathbf{A}_i \mathbf{T}_i \\ \Psi_i &= (\mathbf{T}_i' \mathbf{T}_i)^{-1}\end{aligned}\quad (12)$$

where  $\mathbf{T}_i$  indicates the group-specific  $Q \times Q$  rotation matrix. As opposed to Jennrich (1973), no restrictions are imposed on the diagonal of (any of) the group-specific factor covariance matrices  $\Psi_i$ . Instead, the following restriction is imposed across all groups (see Section 2.3.2.):

$$\frac{1}{I} \sum_{i=1}^I \text{diag}(\Psi_i) = [\mathbf{1}]_Q \quad \text{or} \quad \sum_{i=1}^I \text{diag}(\Psi_i) = I[\mathbf{1}]_Q \quad (13)$$

The differentials of the relations in Equations 12 and 13 are as follows:

$$d\Lambda_i = d\mathbf{A}_i \mathbf{T}_i + \mathbf{A}_i d\mathbf{T}_i ; \quad (14)$$

$$d\Psi_i = -\Psi_i (\mathbf{T}_i' d\mathbf{T}_i + d\mathbf{T}_i' \mathbf{T}_i) \Psi_i ; \quad (15)$$

$$\sum_{i=1}^I \text{diag}(d\Psi_i) = [\mathbf{0}]_Q . \quad (16)$$

Let  $\mathbf{K}_i$  be defined as:

$$\mathbf{K}_i = \mathbf{T}_i^{-1} d\mathbf{T}_i \Psi_i \quad \text{so} \quad d\mathbf{T}_i = \mathbf{T}_i \mathbf{K}_i \Psi_i^{-1} . \quad (17)$$

Equations 14 through 16 then become:

$$d\Lambda_i = d\mathbf{A}_i \mathbf{T}_i + \mathbf{A}_i \mathbf{K}_i \Psi_i^{-1} \quad (18)$$

$$d\Psi_i = -(\mathbf{K}_i + \mathbf{K}_i') \quad (19)$$

$$\sum_{i=1}^I \text{diag}(d\Psi_i) = -\sum_{i=1}^I \text{diag}(\mathbf{K}_i + \mathbf{K}_i') = -\sum_{i=1}^I \text{diag}(2\mathbf{K}_i) = [\mathbf{0}]_Q \quad (20)$$

It follows that  $\sum_{i=1}^I \text{diag}(\mathbf{K}_i) = [\mathbf{0}]_Q$ . Due to these restrictions, the diagonal elements of  $\mathbf{K}_i$  may be

decomposed as follows:

$$\text{diag}(\mathbf{K}_i) = \text{diag}(\tilde{\mathbf{K}}_i) - \frac{1}{I} \sum_{i'=1}^I \text{diag}(\tilde{\mathbf{K}}_{i'}) = \text{diag}(\tilde{\mathbf{K}}_i) - \text{diag}(\bar{\mathbf{K}}). \quad (21)$$

When  $\Lambda_i$  are the optimal rotations for groups  $i = 1, \dots, I$ , the differential in Equation 5 is equal

to zero, thus<sup>6</sup>:

$$\sum_{i=1}^I dR_{mg}(\Lambda_i) = \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} d\lambda_{ijq} = \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{jq}} (d\Lambda_i)_{jq} = 0, \quad (22)$$

where  $(d\Lambda_i)_{jq}$  refers to the element in row  $j$  and column  $q$  of the differential in Equation 18. Since

the optimal rotation restrictions affect the rotated loadings through the rotation matrix  $\mathbf{T}_i$  only, the differential becomes:

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{jq}} (\Lambda_i \mathbf{K}_i \Psi_i^{-1})_{jq} = 0. \quad (23)$$

Since restrictions are imposed (across groups) on the diagonal elements of  $\mathbf{K}_i$ , but not on the offdiagonal elements, we will elaborate Equation 23 for its diagonal and offdiagonal elements separately. To this end, we introduce the matrix  $\mathbf{K}_i^*(u, u)$  which consists of zeros except for the element in row  $u$  and column  $u$ , which is equal to the corresponding element of  $\mathbf{K}_i$ , i.e.,  $k_{iuu}$ . Then Equation 23 is equivalent to requiring:

---

<sup>6</sup> The total differential is the sum of the partial derivatives multiplied by the corresponding differential/infinisimal change.

$$\begin{aligned}
& \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{jq}} (\Lambda_i \mathbf{K}_i^*(u, u) \Psi_i^{-1})_{jq} \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} k_{iuu} \psi_{iuq}^{-1} \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} (\tilde{k}_{iuu} - \bar{k}_{iuu}) \psi_{iuq}^{-1} \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \left[ \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \tilde{k}_{iuu} \psi_{iuq}^{-1} - \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \bar{k}_{iuu} \psi_{iuq}^{-1} \right] \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \tilde{k}_{iuu} \psi_{iuq}^{-1} - \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \bar{k}_{iuu} \psi_{iuq}^{-1} \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \tilde{k}_{iuu} \psi_{iuq}^{-1} - \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \left( \frac{1}{I} \sum_{i'=1}^I \tilde{k}_{i'iuu} \right) \psi_{iuq}^{-1} \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \tilde{k}_{iuu} \psi_{iuq}^{-1} - \frac{1}{I} \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \sum_{i'=1}^I \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \tilde{k}_{i'iuu} \psi_{iuq}^{-1} \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \tilde{k}_{iuu} \psi_{iuq}^{-1} - \frac{1}{I} \sum_{i=1}^I \sum_{i'=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{i'jq}} \lambda_{i'ju} \tilde{k}_{iuu} \psi_{i'uq}^{-1} \\
&= \sum_{i=1}^I \left( \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \tilde{k}_{iuu} \psi_{iuq}^{-1} - \frac{1}{I} \sum_{i'=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{i'jq}} \lambda_{i'ju} \tilde{k}_{iuu} \psi_{i'uq}^{-1} \right) = 0
\end{aligned} \tag{24}$$

Without loss of generality, assume that  $\tilde{k}_{iuu} = 1$ , so that Equation 24 is equivalent to:

$$\begin{aligned}
& \sum_{i=1}^I \left( \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \psi_{iuq}^{-1} - \frac{1}{I} \sum_{i'=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{i'jq}} \lambda_{i'ju} \psi_{i'uq}^{-1} \right) \\
&= \sum_{i=1}^I \left( \sum_{j=1}^J \sum_{q=1}^Q \lambda'_{iju} \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \psi_{iq}^{-1} - \frac{1}{I} \sum_{i'=1}^I \sum_{j=1}^J \sum_{q=1}^Q \lambda'_{i'ju} \frac{\partial R_{mg}}{\partial \lambda_{i'jq}} \psi_{i'qu}^{-1} \right) = 0
\end{aligned} \tag{25}$$

Similarly, by using the matrix  $\mathbf{K}_i^*(u, v)$  where  $u \neq v$  and we can assume that  $k_{iuv} = 1$ , we derive

the following for the offdiagonal elements:

$$\begin{aligned}
& \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{jq}} \left( \Lambda_i \mathbf{K}_i^* (u, v) \Psi_i^{-1} \right)_{jq} \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} k_{iuv} \psi_{iuq}^{-1} \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \lambda_{iju} \psi_{iuq}^{-1} \\
&= \sum_{i=1}^I \sum_{j=1}^J \sum_{q=1}^Q \lambda_{iju} \frac{\partial R_{mg}}{\partial \lambda_{ijq}} \psi_{iqu}^{-1} = 0
\end{aligned} \tag{26}$$

Therefore, the optimal rotation restrictions for each group can be expressed in matrix form as follows:

$$F_{mg} = \Lambda_i' \frac{dR_{mg}}{d\Lambda_i} \Psi_i^{-1} - \text{diag} \left( \frac{1}{I} \sum_{i'=1}^I \Lambda_{i'}' \frac{dR_{mg}}{d\Lambda_{i'}} \Psi_{i'}^{-1} \right) = [\mathbf{0}]_{Q \times Q} \tag{27}$$

where the ‘diag’ operator indicates that the offdiagonal elements are set to zero. Note that, when one prefers to use a reference group with factor variances fixed to one instead of restricting the mean of the factor variances (Section 2.3.2), one can use the restrictions  $\Lambda_i' \frac{dR_{mg}}{d\Lambda_i} \Psi_i^{-1} = [\mathbf{0}]_{Q \times Q}$  for the non-reference groups and the restrictions in Equation 2 (Jennrich, 1973) for the reference group.

### Appendix C

List of items for the Open Sex Role Inventory and what they intend to measure between brackets; 'M' refers to masculinity and 'F' to femininity:

- Q1. I have studied how to win at gambling (M).
- Q2. I have thought about dying my hair (F).
- Q3. I have thrown knives, axes or other sharp things (M).
- Q4. I give people handmade gifts (F).
- Q5. I have day dreamed about saving someone from a burning building (M).
- Q6. I get embarrassed when people read things I have written (F).
- Q7. I have been very interested in historical wars (M).
- Q8. I know the birthdays of my friends (F).
- Q9. I like guns (M).
- Q10. I am happiest when I am in my bed (F).
- Q11. I did not work very hard in school (M).
- Q12. I use lotion on my hands (F).
- Q13. I would prefer a class in mathematics to a class in pottery (M).
- Q14. I dance when I am alone (F).
- Q15. I have thought it would be exciting to be an outlaw (M).
- Q16. When I was a child, I put on fake concerts and plays with my friends (F).
- Q17. I have considered joining the military (M).
- Q18. I get dizzy when I stand up sharply (F).
- Q19. I do not think it is normal to get emotionally upset upon hearing about the deaths of people you did not know (M).
- Q20. I sometimes feel like crying when I get angry (F).
- Q21. I do not remember birthdays (M).
- Q22. I save the letters I get (F).
- Q23. I playfully insult my friends (M).
- Q24. I oppose medical experimentation with animals (F).
- Q25. I could do an impressive amount of push ups (M).
- Q26. I jump up and down in excitement sometimes (F).
- Q27. I think a natural disaster would be kind of exciting (M).
- Q28. I wear a blanket around the house (F).
- Q29. I have burned things up with a magnifying glass (M).
- Q30. I think horoscopes are fun (F).
- Q31. I don't pack much luggage when I travel (M).
- Q32. I have thought about becoming a vegetarian (F).
- Q33. I hate shopping (M).
- Q34. I have kept a personal journal (F).

- Q35. I have taken apart machines just to see how they work (M).
- Q36. I take lots of pictures of my activities (F).
- Q37. I have played a lot of video games (M).
- Q38. I leave nice notes for people now and then (F).
- Q39. I have set fuels, aerosols or other chemicals on fire, just for fun (M).
- Q40. I really like dancing (F).
- Q41. I take stairs two at a time (M).
- Q42. I bake sweets just for myself sometimes (F).
- Q43. I think a natural disaster would be kind of exciting (M).
- Q44. I decorate my things (e.g. stickers on laptop) (F).