Parameter Accuracy in Meta-Analyses of Factor Structures

Timo Gnambs

Leibniz Institute for Educational Trajectories

Thomas Staufenbiel

Osnabrück University

Author Note

Timo Gnambs, Leibniz Institute for Educational Trajectories, Germany; Thomas Staufenbiel, Institute of Psychology, Osnabrück University, Germany;

Correspondence concerning this article should be addressed to Timo Gnambs, Leibniz Institute for Educational Trajectories, Wilhelmsplatz 3, 96047 Bamberg, Germany, Tel.: +49 (0)951 / 863–3420. E-mail: timo.gnambs@lifbi.de

Accepted for publication in *Research Synthesis Methods*.

Abstract

Two new methods for the meta-analysis of factor loadings are introduced and evaluated by Monte Carlo simulations. The direct method pools each factor loading individually, whereas the indirect method synthesizes correlation matrices reproduced from factor loadings. The results of two simulations demonstrated that the accuracy of meta-analytical derived factor loadings is primarily affected by characteristics of the pooled factor structures (e.g., model error, communality) and to a lesser degree by the sample size of the primary studies and the number of included samples. The choice of the meta-analytical method had a minor impact. In general, the indirect method produced somewhat less biased estimates, particularly for small-sample studies. Thus, the indirect method presents a viable alternative for the meta-analysis of factor structures that could also address moderator hypotheses.

Keywords: meta-analysis, factor analysis, factor loadings, factor congruence, bias

Parameter Accuracy in Meta-Analyses of Factor Structures

Psychometric examinations of measurement instruments using meta-analytical methods primarily focus on issues of reliability (e.g., Gnambs, 2014, 2015; Kasten and Freund, 2015) or selected aspects of criterion validity (e.g., Salgado and Táuriz,, 2014; Van Iddekinge et al., 2012); matters of factorial validity are rarely addressed (e.g., Worley et al., 2008). The latter is typically impeded by a lack of appropriate meta-analytic methods for the pooling of published factor loading matrices. Although questions on the factorial configuration of a given instrument can be readily addressed (see Shafer, 2005, 2006) there is no broadly applicable technique for the examination of the magnitude of loadings. Therefore, the present study addresses this lacuna and introduces two meta-analytic techniques to achieve the latter objective. In two Monte Carlo simulations the two approaches are evaluated with regard to the accuracy of the pooled factor loadings.

The Common Factor Model

In common factor theory the observed covariation between a set of *p* manifest variables is assumed to be explained by a smaller set of *q* unobserved common factors. The common factor model can be expressed as follows:

$$\mathbf{P} = \mathbf{\Lambda}\Phi\mathbf{\Lambda}' + \Theta^2 \qquad [1]$$

where $\mathbf{P}$ is a $p \times p$ population correlation matrix, $\mathbf{\Lambda}$ is a p x q factor loadings matrix[1] for q common factors, $\mathbf{\Phi}$ is a q x q matrix of common factor correlations, and $\mathbf{\Theta}^2$ is the p x p diagonal matrix of squared unique factor loadings, that is, the part of each manifest variable unaccounted for by the common factors. The goal of exploratory factor analysis

---

[1] The terms loading matrix and pattern matrix will be used interchangeably throughout the paper.

(EFA) is the identification of the number of common factors and the nature of their relationship with the manifest variables. Thus, EFA tries to estimate $\mathbf{\Lambda}$ given a sample correlation matrix $\mathbf{R}$. More precisely, EFA tries to answer two questions (cf. Meredith, 1993): First, EFA aims at identifying an appropriate loading configuration for $\mathbf{\Lambda}$ that adequately represents the covariation between the p manifest variables. This includes determining the optimal number of common factors q and according to the simple structure criterion for each manifest variable the common factor with the strongest loading. Second, the metric goal of EFA is to provide accurate estimates of the factor loadings in $\mathbf{\Lambda}$. However, a pervasive problem of EFA is that factor analyses of different sample correlations matrices $\mathbf{R}_i$ ($i = 1, \ldots, k$ samples) will rarely identify identical factor pattern matrices $\mathbf{L}_i$, even if the same factor model $\mathbf{\Lambda}$ holds in the population. Sampling error typically distorts the $\mathbf{L}_i$ and, thus, leads to biased estimates of $\mathbf{\Lambda}$ for different samples (see MacCallum and Tucker, 1991, for an in-depth discussion of sampling error in EFA). As a potential solution to this problem it has been suggested to identify the true common factor model for a set of manifest variables using meta-analytic methods (e.g., Becker, 1996; Kaiser et al., 1971; Shafer, 2005, 2006).

Meta-Analytic Approaches for Factor Structures

So far, there are three major avenues for the meta-analytical examination of factor structures: The first approach involves pairwise comparisons of the k sample pattern matrices $\mathbf{L}_i$ that have been accrued from primary studies (Bushman et al., 1991). After rotation to maximal similarity the loading matrices are compared using the coefficient of congruence (Tucker, 1951). A strength of this approach is the sole reliance on data typically reported for EFAs in published research articles which, thus, are likely to be available for meta-analytic integration. However, this advantage is typically outweighed by a major limitation: The approach only provides pairwise comparisons between factor loadings patterns and no overall index of similarity. Moreover, it does not provide an estimate of $\mathbf{\Lambda}$ but only a form of fit index reflecting the heterogeneity of results. The second approach is a form of a count technique introduced by Shafer (2005, 2006) which involves counting the number of times two manifest variables exhibited their strongest loading on the same common factor across all k samples. Using these co-occurrences the factor configuration of $\mathbf{\Lambda}$ is determined. An advantage of Shafer's method is the use of loading structures from EFAs that are typically reported in primary studies. However, a weakness is that it is limited to the examination of the loading configuration (i.e. to determine which items load on the same or different factors); but it does not acknowledge the magnitude of the reported factor loadings. The third approach adopts a two-step strategy: Correlations between items of a scale are aggregated using conventional uni- or multivariate meta-analytical techniques. Subsequently, the pooled correlation matrix is analyzed using common exploratory or confirmatory factor analytical methods (cf. Becker, 1996; Cheung, 2014; Cheung and Chan, 2005). Although this approach can address configural as well as metric aspects in EFA, it is seldom used in practice. Except for some rare cases (see for example, Gnambs, 2013) item-level correlation matrices are not typically reported in primary studies and, thus, the meta-analysis of the sample correlation matrices does not represent a widely applicable option for the examination of factor structures.

Proposals for Alternative Meta-Analytic Approaches

Although factor configurations can be examined with existing meta-analytic techniques (Shafer, 2005, 2006), there is no broadly applicable method for the analyses of metric information, that is, the magnitude of the loading pattern. Therefore, two avenues are introduced that could be used to address the latter issue (the major steps involved in these methods are also briefly summarized in step 4 of Table 1):

The first approach (*direct* method) refers to the pooling of the reported factor loading matrices. Since in EFA an infinite number of rotated solutions explain the data equally well, the factor pattern matrix is not unique. Hence, before aggregating over the *k* sample factor loadings matrices $\mathbf{L}_i$ they are first rotated to achieve maximal similarity. Because the true population factor model $\mathbf{\Lambda}$ is unknown, it is suggested to rotate all $\mathbf{L}_i$ towards an idealized target structure **T** specifying 1s for all hypothesized salient loadings and 0s for non-salient loadings. Thus, **T** reflects the hypothesized factor configuration. Of course, this assumes that **T** is derived *a priori*, typically based on theoretical considerations and the results of empirical studies. In case of ambiguous loadings patterns for specific items it is also possible to use a partially specified target for **T** (Browne, 1972). Subsequently, the rotated sample factor loadings $\tilde{\mathbf{L}}_i$ are pooled into $\bar{\mathbf{L}}$ as an estimate of $\mathbf{\Lambda}$.

The second approach (*indirect* method) is based on the pooling of the reproduced correlation matrices. Following the common factor model the implied (or reproduced) correlation matrix $\mathbf{R}^*$ of a given sample factor pattern matrix **L** and the respective sample factor correlation matrix **F** is given by

$$\mathbf{R}^* = \mathbf{LFL}'. \qquad [2]$$

After aggregating the *k* reproduced correlation matrices $\mathbf{R}_i^*$ the pooled correlation matrix $\bar{\mathbf{R}}^*$ is subjected to a conventional exploratory factor analysis that yields $\bar{\mathbf{L}}$ as an estimate of $\mathbf{\Lambda}$. Thus, the indirect approach can be conceived as an extension of Becker's

(1996) two-step strategy, albeit based on the reproduced instead of the sample correlation matrices. Although both approaches achieve the same goal of providing an estimate of $\mathbf{\Lambda}$ the indirect approach is more flexible because it does not assume a priori knowledge of a specific factor configuration and, thus, could even be used to aggregate loading patterns with a different number of factors. Moreover, it could also be extended to the examination of competing measurement models (e.g., using confirmatory factor analyses), similar to meta-analytic structural equation modelling (MASEM; cf. Cheung, 2014, 2015; Cheung and Chan, 2005, 2009).

Both approaches, the direct and the indirect method, require the pooling of effect sizes (either factor loadings or correlations) across multiple samples. This can be achieved by adopting univariate techniques that individually pool each of the *s* unique effects[2] or multivariate techniques that also acknowledge within-sample dependencies between effects (cf. Cheung, 2013). A general framework for univariate and multivariate meta-analyses represents the generalized least squares (GLS) technique (Becker, 1992): Let **y** be the $k{\cdot}s$ x 1 column vector of the *k* stacked *s* x 1 vectors of the sample's observed effects sizes for the meta-analyses (i.e. factor loadings for the direct method and correlations for the indirect method). Then, the GLS estimator for the population effect sizes $\boldsymbol{\beta}$ is also a *s* x 1 column vector given as

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \left(\sum_{i=1}^{k}\mathbf{V}_i^{-1}\right)^{-1}\sum_{i=1}^{k}\mathbf{V}_i^{-1}\mathbf{y}_i \qquad [3]$$

---

[2] In the direct approach $s = p{\cdot}q$ factor loadings, whereas in the indirect approach $s = p{\cdot}(p\text{-}1) / 2$ correlations.

where **X** consists of *k* stacked *s* x *s* identity matrices, and **V** is a *k*·*s* x *k*·*s* blockwise diagonal matrix consisting of each samples *s* x *s* variance-covariance matrix of the effect sizes arrayed along the diagonal (Becker, 1992; Hafdahl, 2007). Thus, the GLS estimation of **β** corresponds to the calculation of the variance-weighted means of the observed effect sizes **y** (Hedges and Vevea, 1998). In meta-analyses **V** is assumed to be known. If all off-diagonal elements in **V** are set to zero this corresponds to a conventional univariate meta-analysis for each effect size. In contrast, if dependencies between the effect sizes within a sample are acknowledged by calculating the covariances in **V** (cf. Furlow and Beretvas, 2005), the model corresponds to a multivariate meta-analysis (Becker, 2002; Cheung, 2013).

Interestingly, under a fixed-effects model that assumes homogenous population effect sizes across samples the GLS estimator in [3] reduces to

$$\hat{\boldsymbol{\beta}} = \sum_{i=1}^{k} n_i \mathbf{y}_i \Big/ \sum_{i=1}^{k} n_i \qquad [4]$$

where n_i is the size of the *i*th sample (Hafdahl, 2007). Thus, in the fixed-effects case within-study dependencies between effect sizes do not affect the estimation of the population effect and the GLS estimator for **β** is simply the sample-size weighted mean of effects. Therefore, for the direct method the pooled factor loadings can be estimated as sample-size weighted mean loadings across the *k* samples and for the indirect method the pooled correlation matrices are derived as sample-size weighted mean correlations.

Random-effects model would require estimates of the variance-covariance matrix **V**; that is, the sampling variances of the reported factor loadings for the direct method or the sampling variances of the reproduced correlations for the indirect method. However, analytical solutions for the calculation of these values based on reported factor loadings alone are not available. Therefore, a potential approximation for **V** would be the use of respective formulas developed for correlation coefficients (see Hedges, 1983; Olkin and Finn, 1996). But in practice accurate estimation of **V** might be unnecessary because it does not improve the pooled estimates. Simulation studies indicate that sample-size weights result in less biased estimates of correlation coefficients than inverse variance weights (Brannick et al., 2011). Moreover, taking the covariances in **V** into account does not improve point-estimates of the pooled correlations and subsequent structural models (Prevost et al., 2007). Therefore, given the complexities involved in estimating **V** and the uncertain advantages the present study will focus on the fixed-effects case.

Comparison of Approaches

The two proposed meta-analytic methods have several advantages as compared to available techniques (see Table 2). In contrast to MASEM, the direct and indirect method put less demands on the primary studies because they only require information on the factor loadings instead of the item-level correlations. While the latter are rarely reported in primary studies, the former are routinely presented in EFA studies. Thus, for meta-analyses adopting one of the new approaches it is likely that a larger number of effects would be available, thus, potentially increasing the precision of the pooled estimates. With regard to the research questions that can be addressed, the direct and the indirect method show two central differences. First, metric issues (i.e. concerning the magnitude of factor loadings) can be similarly examined using the direct and the indirect method. In contrast, the analyses of configural questions (i.e. regarding the appropriate number of underlying factors) are limited to the indirect method. Second, in the context of the direct method moderator hypotheses can only be addressed for each factor loading separately using, for example, traditional meta-regressive techniques (see Viechtbauer et al., 2014, for an overview). In contrast, moderators in the indirect approach could either focus on individual factor loadings by examining gradients of factor loadings (Gnambs, 2013) or examine moderator effects for the entire factor structure within the established framework of multi-group invariance testing (Davidov et al., 2014). Thus, in many ways the indirect method is similar to MASEM and permits addressing similar research objectives (see Table 2). However, the use of factor loadings (instead of item-level correlations) as raw data extracted from the primary studies makes the indirect method potentially a better choice for factor analytic studies.

Hypotheses for the Present Studies

Before adopting the direct or the indirect method for the meta-analysis of factor structures their performance with regard to established meta-analytic techniques needs to be demonstrated. Therefore, this study examined the accuracy of the pooled factor loadings derived from the direct and indirect method in relation to Becker's (1996) two-step strategy that assumes availability of the item-level correlation matrices. In particular, the study focuses on the magnitude of the factor loadings and potential differences in loading biases between the three meta-analytic methods. Moreover, the study examined several specific hypotheses with respect to variables that might affect the accuracy of $\bar{\mathbf{L}}$ as an estimate of $\mathbf{\Lambda}$.

Hypothesis 1: Typically, sampling error distorts estimates of factor loadings in EFA. Several studies demonstrated that the influence of sampling error can be reduced by increasing the sample size in EFA (e.g., Guadagnoli and Velicier, 1989; Hogarty et al., 2005; MacCallum et al., 1999; MacCallum et al., 2001; Preacher and MacCallum, 2002). Although the effect of the number of participants on EFA results depends on various additional conditions such as the degree of overdetermination or the communality of the items larger sample sizes typically tend to result in less biased factor loadings. Therefore, it is assumed that there would be a positive effect of the average sample size on the accuracy of $\bar{\mathbf{L}}$ as an estimate of $\mathbf{\Lambda}$.

Hypothesis 2. In meta-analyses sampling error of individual samples is minimized by pooling effects across different samples. Increases in the number of effect sizes in a meta-analyses typically yields better control of sampling error and more accurate recovery of the population effects (e.g., Field, 200l, 2005; Hafdahl and Williams, 2009). Moreover, research on MASEM (e.g., Cheung and Chan, 2005) indicated that parameter accuracy of structural parameters tends to increase with the number of samples in the meta-analysis. Therefore, similar effects are expected with regard to the bias in $\bar{\mathbf{L}}$: There would be a positive effect of the number of samples on the accuracy of $\bar{\mathbf{L}}$ as an estimate of $\mathbf{\Lambda}$.

Hypothesis 3. Simulation studies indicate that factor loading biases in EFA are strongly influenced by the items' communalities, that is, the proportion of variance accounted for by the major common factors (e.g., Hogarty et al., 2005; MacCallum et al., 1999, 2001; de Winter et al., 2009). If communalities are high EFA results in less biased factor loadings. Therefore, it is expected that there would be a positive effect of the average communality of items on the accuracy of $\bar{\mathbf{L}}$ as an estimate of $\mathbf{\Lambda}$.

Hypothesis 4. Model error refers to the lack of fit of the population factor model specified in [1] to the data. The common factor model described in [1] represents an idealized situation that rarely holds in practice. Rather, a number of unobserved minor common factors **W** also contribute to **P** (MacCallum et al., 2001):

$$\mathbf{P} = \mathbf{\Lambda}\Phi\mathbf{\Lambda}' + \Theta^2 + \mathbf{W}\mathbf{W}' . \qquad [5]$$

Although the minor common factors cannot be identified empirically using EFA their presence distorts estimates of $\mathbf{\Lambda}$ (even in the absence of sampling error) and, thus, influences the factor loading estimates (Briggs and MacCallum, 2003; Preacher and MacCallum, 2002). Therefore, it is expected that there would be a negative effect of the degree of model error on the accuracy of $\bar{\mathbf{L}}$ as an estimate of $\mathbf{\Lambda}$.

Hypothesis 5. The determination of the number of factors that underlie the data is sometimes a difficult task. As a consequence overextraction (extracting too many factors) or underextraction (extracting too few factors) can result. In general, underextraction is seen as a more serious problem than overextraction (Fava and Velicer, 1996; Wood, et al., 1996). In case of the indirect method underextraction could be in particular disadvantageous because the information for reproducing the correlation matrix is diminished. Therefore, it is expected that for an increasing number of included studies in which underextraction occurred the accuracy of $\bar{\mathbf{L}}$ as an estimate of $\mathbf{\Lambda}$ would reduce.

Hypothesis 6. The situation that primary studies with too few extracted factors exist can be handled in two different ways. The samples in which underextraction occurred can be included or excluded from the analysis. Whereas the latter approach reduces the number of samples in the meta-analyses it considers only valid factor structures. The former approach does not affect the number of samples but includes potentially biased factor structures. Whether inclusion or exclusion of studies in which underextraction occurred has a stronger negative effect on the precision with which $\bar{\mathbf{L}}$ estimates $\mathbf{\Lambda}$ cannot be derived on an *a priori* basis.

Hypothesis 7. Primary studies frequently neglect to report all required information for meta-analytic integration. In particular, some authors only report loadings exceeding an a priori set cutoff value (e.g., .3 or .4) and neglect to report smaller factor loadings that are considered insubstantial. Just like in underextraction the information available to reproduce the correlation matrix is reduced. Moreover, there is a systematic omission of smaller loadings which could distort the pooled factor loadings matrices. Therefore, it is expected that there would be a negative effect of the degree of reporting omissions on the accuracy of $\bar{\mathbf{L}}$ as an estimate of $\mathbf{\Lambda}$.

These hypotheses were examined in two independent Monte Carlo simulations. The first simulation focused on differences between the two proposed meta-analytic methods and examined Hypotheses 1 to 4. Moreover, the results for the indirect and direct method were compared to an established meta-analytic technique for the pooling of item-level correlation matrices (see Becker, 1996). A second simulation study investigated the performance of the indirect approach under additional conditions and focused on Hypotheses 5 to 7.

Study I

A Monte-Carlo simulation examined the parameter accuracy of factor structures derived from direct and indirect meta-analyses of loading matrices. In addition, the experimental design evaluated two aspects of the examined population factor model (i.e. communality and model error) and two central influences on the meta-analytic results (i.e. sample size of primary studies and number of included studies).

Method

A schematic representation of the simulation procedure is summarized in Table 1. The entire simulation included six steps: Whereas step 4 refers to the focal meta-analytic procedures the remaining steps set up the current simulation (from data generation to the calculation of the accuracy indices).

Data generation. A large number of sample correlation matrices were generated from known population correlation matrices that approximately represented a common factor model (step 1 in Table 1). The population correlation matrices were created using the Tucker, Koopman, and Linn (1969; see also MacCallum and Tucker, 1991) procedure that defines each manifest variable as a linear combination of three types of latent factors: major common factors, minor common factors, and unique factors as given in [5]. Minor common factors were included to create more realistic correlation matrices that *approximately* represent the major common factor structure because exact fit almost never holds in practice (cf. MacCallum et al., 2001). The factor model used to generate the population correlation matrices included $p = 10$ items loading on $q = 2$ major common factors and 100 minor common factors. The two major factors were correlated at $\varphi_{12} = .30$, whereas the minor common factors were uncorrelated among each other and also uncorrelated to the major factors. Five items had salient loadings on the first major factor, whereas the remaining variables loaded strongest on the second major factor. The communalities for each manifest variable and, consequently, factor loadings were varied according to the experimental conditions (see below). Non-salient loadings on the major factors for each variable were specified to account for about 10% of the item's communality. The population correlation matrices were simulated in *R* using the algorithms outlined in Hong (1999).

In step 2 (see Table 1) item responses were generated from a multivariate normal distribution with the known population correlation structure using the eigen-decomposition method (Rizzo, 2008). The randomly generated item responses for each simulated sample were subjected to an exploratory maximum likelihood factor analysis that extracted two factors and obliquely rotated the pattern matrix using the oblimin criterion with $\delta = 0$ (step 3 in Table 1). Factor analyses that yielded Heywood cases (i.e. communalities greater than 1) were discarded and replaced with a valid factor solution.[3] The resulting sample factor loading matrices represented the effect sizes for the meta-analyses.

Experimental design. Five design factors were manipulated in this study to evaluate their impact on the simulation results. The first design factor examined the *meta-analytical method*: (a) In the direct method for the meta-analysis of factor structures, the factor loading matrices were aggregated using the fixed-effects GLS estimator as described above in [4]. Before synthesizing the loadings each sample factor structure was rotated

---

[3] Because meta-analyses rely on published factor solutions, it seemed unlikely that improper factor solutions with Heywood cases would be available and, thus, could be incorporated into meta-analyses of factor loadings. Moreover, previous research indicated that screening results and replacing improper with valid factor solutions does not distort Monte Carlo simulations of factor analyses (MacCallum et al., 1999, 2001).

toward an idealized target structure **T** specifying 1s for all hypothesized salient loadings and 0s for hypothesized non-salient loadings (cf. step 4a in Table 1). The orthogonal Procrustes rotation applied here minimizes the squared differences between **T** and the loadings matrix (Schönemann, 1966). (b) In the indirect method for the meta-analysis of factor structures, the reproduced sample correlation matrices were calculated for each sample (step 4a in Table 1) and subsequently pooled using the same meta-analytical procedure as in the direct method. Factor loadings were derived by subjecting the pooled correlation matrix to an exploratory maximum likelihood factor analysis with oblimin rotation using $\delta = 0$ (step 4c in Table 1). (c) As a control condition, the third meta-analytic method aggregated the sample correlation matrices generated in step 2 (Table 1). This condition represents the rare situation where item-level correlation matrices would be available for meta-analysis (cf. Becker, 1996). These correlation matrices were pooled using the fixed-effects GLS estimator given in [4] (step 4a in Table 1); subsequently, the pooled correlation matrix was subjected to an exploratory maximum likelihood factor analysis with oblimin rotation (step 4b).

The second design factor varied the *proportion of item variance accounted for* by the two major common factors. In line with previous research (e.g., MacCallum et al., 1999, 2001), we examined three levels: ‘low’ with values between .2 and .4; ‘wide’ with values varying between .2 and .8; and, ‘high’ with values falling between .6 and .8. The respective major common factor pattern matrices for the three conditions are presented in Appendix A. Thus, depending on the experimental condition a different population correlation matrix was generated (see Appendix B) that specified the respective proportion of accounted variance for each manifest variable. Because of the inclusion of minor common factors these values are not entirely equivalent to the communality h^2 of an item. However, communalities vary with the proportion of accounted variance. Therefore, this experimental condition will be referred to as low, wide, or high communalities, henceforward.

The third design factor examined the level of *model error* that indicates the lack of fit of the major common factor model. Model error can be introduced by manipulating the proportion of variance for each manifest variable accounted for by the minor factors in [5] (Hong, 1999; Tucker et al., 1969). We acknowledged two levels of model error that accounted for either 10% (small error) or 20% (large error) of the item variance. These two conditions correspond to root mean squared error of approximations (RMSEA) for the major common factor model of .03 or .09. According to prevalent interpretations (Schermelleh-Engel et al., 2003) values or RMSEA < .05 indicate good fit whereas RMSEA > .08 are considered not acceptable. Thus, depending on the experimental condition a different population correlation matrix with either small or large model error was generated (see Appendix B).

The fourth design factor varied the *number of samples* included in the meta-analysis. Typical, meta-analyses of factor structures include between 20 and 30 studies (e.g., Shafer, 2005, 2006; Worley et al., 2008) and rarely exceed 50 studies (e.g., Huang and Dong, 2012). Therefore, the simulation included values representing small, moderate, and large number of samples included in a meta-analysis: 10, 30, and 50.

The fifth design factor varied the *average sample size* in each meta-analysis. Sample sizes were randomly drawn from a normal distribution with a mean of the average sample size and a standard deviation of one fourth of the average sample size (cf. Cheung and

Chan, 2005). The average sample sizes (and their standard deviations) considered in this simulation were 60 (15), 200 (50), and 400 (100). These values represent small, medium, and large sample sizes for exploratory factor analyses (MacCallum et al., 2001). Because sample sizes were randomly generated and, thus, could theoretically become negative, the minimum sample size considered in this study was 30 (i.e. randomly generated sample sizes falling below this threshold were truncated to 30).

Whereas the method of aggregation represented a within-replication factor (i.e. the three meta-analytic methods were applied to the same simulated samples), the remaining design factors were manipulated between replications.

Evaluation criteria. The pooled factor structures were evaluated by conducting one thousand replications in each experimental condition (step 6 in Table 1). The parameter estimates were evaluated by three indices comparing the aggregated factor loading matrices to the population factor loading matrices: coefficient of alienation, (relative) bias, and (relative) mean squared deviation. Before calculating these indices the aggregated factor loading matrices $\bar{\mathbf{L}}$ were rotated to an optimal fit with the population factor matrix **T** using orthogonal Procrustes rotation yielding $\hat{\mathbf{\Lambda}}$ (step 5 in Table 1). To derive at the orthogonal Procrustes rotation of $\bar{\mathbf{L}}$ we use a singular value decomposition of matrix $\mathbf{T}'\bar{\mathbf{L}}$

$$\mathbf{T}'\bar{\mathbf{L}} = \mathbf{P\Omega Q}' \quad [6]$$

where $\mathbf{P}'\mathbf{P} = \mathbf{Q}'\mathbf{Q} = \mathbf{I}$ and $\mathbf{\Omega}$ is the diagonal matrix with the singular values. Then

$$\hat{\mathbf{\Lambda}} = \bar{\mathbf{L}}\mathbf{QP}'. \quad [7]$$

The coefficient of alienation (*K*) indicates the dissimilarity of factor structures with values approaching 0 representing more similar factor structures. In this study, *K* was derived for each replication on basis of the coefficient of congruence r_c for complete factor loading matrices (Gebhardt, 1968) which is an extension of the more popular coefficient for single factor congruence (Tucker, 1951). r_c is given by

$$r_c = \frac{\sum_{l=1}^{p}\sum_{m=1}^{q}\hat{\lambda}_{lm}\lambda_{lm}}{\sqrt{\sum_{l=1}^{p}\sum_{m=1}^{q}\hat{\lambda}_{lm}^{2} \cdot \sum_{l=1}^{p}\sum_{m=1}^{q}\lambda_{lm}^{2}}}. \qquad [8]$$

Because r_c is typically very close to its upper bound 1, to allow for a better discrimination it was transformed into the coefficient of alienation (cf. Borg and Leutner, 1985) by

$$K = \sqrt{1 - r_c^2} \qquad [9]$$

which reflects the degree of *in*congruence[4]. A summary statistic was calculated by averaging K across replications.

The accuracy of the factor loadings in $\hat{\mathbf{\Lambda}}$ as estimators of $\mathbf{\Lambda}$ was examined using a measure of absolute bias defined as $B(\hat{\mathbf{\Lambda}}) = \mathrm{E}(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda})$. To facilitate interpretations we did not examine each element in $\hat{\mathbf{\Lambda}}$ separately[5], but calculated the mean bias for the entire factor loading matrix yielding

---

[4] All analyses were also repeated using r_c. Because respective results closely mirrored those for K, they are not presented.

[5] We also examined bias for each parameter in $\hat{\mathbf{\Lambda}}$ individually. However, since these analyses did not lead to markedly different conclusions but replicated the pattern of effects of the reported results, the respective analyses are not presented here.

$$B(\hat{\mathbf{\Lambda}}) = E\left(\frac{\sum_{l=1}^{p}\sum_{m=1}^{q}(\hat{\lambda}_{lm} - \lambda_{lm})}{p \cdot q} \right) \qquad [10]$$

(see MacCallum et al., 2001, and Myers et al., 2015, for a similar index). Similarly, we calculated the average relative bias (Hoogland and Boomsma, 1998) for the entire factor loading matrix as

$$B_r(\hat{\mathbf{\Lambda}}) = E\left(\frac{\sum_{l=1}^{p}\sum_{m=1}^{q}(\hat{\lambda}_{lm} - \lambda_{lm}) / |\lambda_{lm}| \cdot 100}{p \cdot q} \right) \qquad [11]$$

In line with prevalent practice relative biases below 10% (Forero et al., 2009) or 5% (Hoogland and Boomsma, 1998) are considered insubstantial, whereas values above 20% are deemed unacceptable.

Finally, the efficiency of the new meta-analytic methods is reported in terms of the mean squared error (*MSE*) which is given as $MSE(\hat{\mathbf{\Lambda}}) = \mathrm{E}\left[(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda})^2\right]$. Because $MSE(\hat{\mathbf{\Lambda}})$ can be simplified to $MSE(\hat{\mathbf{\Lambda}}) = B(\hat{\mathbf{\Lambda}})^2 + VAR(\hat{\mathbf{\Lambda}})$ a more biased estimator might be more efficient if it yields a considerable smaller variance (Mood et al., 1973). Again, *MSE* was calculated as the mean error for the entire factor loading matrix (Velicer and Fava, 1998):

$$MSE(\hat{\mathbf{\Lambda}}) = E\left(\frac{\sum_{l=1}^{p}\sum_{m=1}^{q}(\hat{\lambda}_{lm} - \lambda_{lm})^2}{p \cdot q} \right). \qquad [12]$$

Moreover, two estimators can be compared based on their relative efficiency (Hafdahl and Williams, 2009) using $\left(MSE(\hat{\mathbf{\Lambda}}_1) / MSE(\hat{\mathbf{\Lambda}}_2) - 1\right) \cdot 100$.

In addition to evaluating the entire factor pattern, *B* and *MSE* were also calculated separately for salient and non-salient factor loadings using [10], [11], and [12] on a subset of loadings in $\hat{\mathbf{\Lambda}}$.

Results

The effects of the experimental manipulations on the three evaluation criteria were analyzed using 3 (method of aggregation: direct, indirect, control) x 3 (communality: low, wide, high) x 2 (model error: small, large) x 3 (number of samples: 10, 30, 50) x 3 (average sample size: 60, 200, 400) analyses of variance (ANOVA). Given the skewed distribution of *K* the respective analysis was performed on the Fisher's *Z* transformed coefficients of alienation. Because of the large number of replications in each condition the *F*-tests had excessive power and yielded significant results even for negligible effects. Therefore, the interpretations of these results focus on generalized eta squared η_G^2 as effect size (Bakeman, 2005; Olejnik and Algina, 2003). η_G^2 indicates the proportion of explained variance by a main effect or interaction after eliminating the effects of the other experimental factors. In line with prevalent recommendations (Ferguson, 2009), η_G^2 exceeding .04 are considered practically relevant effects. Post-hoc analyses for within-replication factors are based on paired-sample *t*-tests and Hedges (1981) *g* as effect size (see Dunlap, Cortina, Vaslow and Burke, 1996, for *g* in repeated measurement designs).

Similarity of factor loading pattern. As summarized in Table 3, factor incongruence was most strongly influenced by the level of communality (η_G^2 = .91) and model error (η_G^2 = .83). Factor patterns with larger communalities and less error resulted in smaller *K* (see Figure 1). Moreover, the interaction (η_G^2 = .80) indicated that incongruence was largest for factor structures containing low communalities and large error at the same time. Thus, characteristics of the population factor model influenced the similarity between the meta-analytic and population factor structure most strongly. In addition, the average sample size (η_G^2 = .22) and the number of samples included in the meta-analysis (η_G^2 = .14) had non-negligible impact on factor incongruence. Factor loadings derived from larger samples and increases in the number of primary studies were associated with smaller *K*. To some degree the accuracy of small-samples meta-analyses could be improved by increasing the number of included samples, as indicated by the respective interaction (η_G^2 = .08). In contrast, the number of samples had only a small impact on meta-analyses with average sample sizes of 200 or 400. Moreover, these effects were most pronounced for the low communality conditions (see left column in Figure 1), as indicated by the interaction between communality and sample size (η_G^2 = .12) or number of samples (η_G^2 = .08). With regard to the adopted meta-analytic approach, the simulation identified a small effect (η_G^2 = .07). Post-hoc analyses showed that this effect was primarily due to the direct method resulting in larger incongruence than the indirect method, $t(53{,}999) = 156.74$, $p < .001$, g = 0.11, and the control condition, $t(53{,}999) = 140.43$, $p < .001$, g = 0.12. In contrast, the indirect method produced similar results as the control condition, $t(53{,}999) = 11.58$, $p < .001$, g = 0.00. Particularly, for small sample meta-analyses the indirect method for the meta-analysis of factor structures slightly outperformed the direct approach (see top row of

Figure 2). All higher-order interactions had negligible effects (η^2_G < .04) and are thus not reported in Table 3.

Accuracy of estimators. Systematic over- and underestimation of factor loadings in the form of absolute bias *B* did not reveal pronounced effects with regard to the three estimators (see Table 3). Although bias tended to increase with the level of communality (η^2_G = .40), all three meta-analytical methods tended to overestimate the population factor loadings slightly (η^2_G = .10). In particular, the direct method resulted in slightly larger bias than the indirect method, *t*(53,999) = 164.19, *p* < .001, *g* = 0.55, and the control condition, *t*(53,999) = 165.68, *p* < .001, *g* = 0.44. The difference between the indirect method and the control condition was considerably smaller, *t*(53,999) = 110.92, *p* < .001, *g* = 0.15, with slightly larger bias for the indirect method. Overall, the relative bias B_r was rather large for all three estimators and fell at about 12% for the direct method as compared to about 10% for the indirect method and the control condition.

Separate analyses conducted for the salient and non-salient factor loadings (see Table S1 in the online supplement) indicated that the observed differences in *B* for the three meta-analytical methods were limited to the non-salient loadings: Whereas the absolute bias in non-salient loadings was affected by the meta-analytic method (η^2_G = .19)—being largest for the indirect method as compared to the direct method, *t*(53,999) = 171.85, *p* < .001, *g* = 0.50, or the control condition, *t*(53,999) = 166.97, *p* < .001, *g* = 0.50, but negligible for the direct method as compared to the control condition, *t*(53,999) = 12.59, *p* < .001, *g* = 0.01—respective analyses for salient loadings did not identify an effect (η^2_G = .02). As a consequence the relative bias B_r was negligible for salient factor loadings (less than 2% for all three estimators), whereas the relative bias was rather severe for non-salient loadings (25% for the indirect method versus 22% for the indirect method and the control condition). Thus, none of the available meta-analytic method was able to adequately estimate the magnitude of non-salient factor loadings.

Efficiency of estimators. The pattern of results for the mean squared error (see Table 3) is similar to that for *K*. Whereas the level of communality (η_G^2 = .54), model error (η_G^2 = .77) and the interaction of both factors (η_G^2 = .70) had the strongest effects on *MSE* (see Figure 3 for the respective root mean squared errors[6]), the impact of the average sample size (η_G^2 = .21) and the number of studies (η_G^2 = .13) was considerably smaller. Again, the method of aggregation had a small effect on the simulation results (η_G^2 = .05). Post-hoc analyses indicated that the direct method showed significantly larger *MSE* than the indirect method, $t(53{,}999) = 149.22$, $p < .001$, $g = 0.15$, $p < .001$, and the control condition, $t(53{,}999) = 126.88$, $p < .001$, $g = 0.15$. In contrast, the indirect method did not differ from the control condition, $t(53{,}999) = 0.97$, $p = .33$, $g = 0.00$. Overall, the indirect method resulted in slightly smaller *MSE*, particularly for small-sample meta-analyses, than the direct method (see bottom row in Figure 2). In terms of relative efficiency, the control condition was about 9% more efficient than the direct method, whereas it showed no superior efficiency (0%) in comparison to the indirect method. Similarly to bias *B*, the observed differences in *MSE* for the three meta-analytical methods were limited to the non-

---

[6] For illustrative purposes *MSE* was transformed into the root mean squared error as $RMSE(\hat{\mathbf{\Lambda}}) = \sqrt{MSE(\hat{\mathbf{\Lambda}})}$

.

salient loadings, η^2_G = .10 (see bottom row in Figure S1 and Table S1 in the online supplement). Aggregation of the salient factor loadings was not affected by the adopted meta-analytical approach, η^2_G = .00 (top row in Figure S1). Again, post-hoc analyses showed that these differences for the non-salient loadings were due to the direct method resulting in larger *MSE* than the indirect method, $t(53{,}999) = 167.24$, $p < .001$, $g = 0.40$, and the control condition, $t(53{,}999) = 152.82$, $p < .001$, $g = 0.39$. In contrast, the indirect method was comparable to the control condition, $t(53{,}999) = 11.48$, $p < .001$, $g = 0.01$. These results translated into relative efficiencies of 14% and 0% for the control condition over the direct and the indirect method, respectively.

Auxiliary analyses. The previous results demonstrated rather small differences between the indirect method that calculated the implied correlation matrix $\mathbf{R}^*$ from the reported factor loadings and the control condition that used the original sample correlation matrix $\mathbf{R}$. Potential differences between the two meta-analytic techniques would be due to a loss of information resulting from conducting an EFA before pooling the correlations. Accuracy and efficiency indices calculated between the two correlation matrices $\mathbf{R}$ and $\mathbf{R}^*$ (see Figure 4) indicated that the accuracy of the reproduced correlations are strongly affected by the communalities of the items. The coefficient of alienation K showed that for larger communalities the sample and reproduced correlation matrices differed by less than 1 percent. In contrast, for small communality matrices respective differences increased up to 6 percent. Thus, potential advantages of meta-analyses of factor structures using the item-level correlation matrices (Becker, 1996; Cheung, 2014; Cheung and Chan, 2005) would be limited to ill-defined factor structures that are dominated by items with small communalities.

Study II

The second simulation study further investigates the accuracy of the indirect method under two conditions that occur frequently in practice: Too few factors are extracted or the loading matrix is not fully reported. Therefore, the Monte Carlo simulation examined the effect of underextraction (Hypotheses 5 and 6) and reporting omissions (Hypothesis 7) on the parameter accuracy of meta-analytical estimates using the indirect method.

Method

Data generation. Sample correlation matrices were simulated adopting a similar procedure as in the previous study. However, sample data was derived from a single population correlation matrix that reflected common conditions for EFA in practice: It resembled the wide communality condition with small model error in the previous study; that is, the percentage of variance accounted for by the two major common factors fell between .2 and .8 and the fit of the major factor model was RMSEA = .03. Moreover, the correlation between the two major common factors was set at $\varphi_{12} = .49$ to increase the likelihood of underextraction in the experimental design (see below) The average sample size and the number of samples included in the simulated meta-analyses were held constant at values of 200 and 30, respectively, that reflect typical values for meta-analyses of factor structures (cf. Shafer, 2005, 2006; Worley et al., 2008). The remaining conditions of the data generation procedure were identical to the previous simulation.

Experimental design. Three design factors were manipulated in this study to evaluate their impact on the simulation results. The first design factor varied the *percentage of underextraction* included in the meta-analysis. This manipulation was implemented by calculating ω_h for each randomly generated sample (McDonald, 1999) which quantifies the proportion of test variance captured by a single factor. For samples with ω_h falling either below 1.00, .50, or .45 the correct factor analytical model was specified; that is, the exploratory factor analysis extracted two factors. In contrast, for samples with ω_h exceeding these thresholds an invalid factor analytic model was specified; that is, the exploratory factor analysis extracted a single factor. As a consequence, the percentage of underextraction fell at about 0%, 10%, and 32% for the three examined conditions. The respective thresholds for ω_h were derived on a trial-and-error basis to approximately result in the specified percentages of underextraction.

The second design factor examined the *method of handling underextraction* by either including or excluding samples with one extracted factor.

The third design factor studied the effect of *reporting omissions* when primary studies only report strong loadings exceeding an a priori set criterion (e.g., .3 or .4) but neglect to report smaller factor loadings. Factor loadings falling below .3 are frequently viewed as negligible (e.g., Fabrigar et al., 1999; Henson & Roberts, 2006) and, thus, might be systematically missing for meta-analyses. This effect was studied by substituting all loadings in a sample falling below .3 as missing and assuming loadings of 0 in the meta-analysis. We considered three conditions that specified 0%, 10%, or 25% of samples included in a meta-analysis as exhibiting reporting omissions.

Whereas the method of handling underextraction represented a within-replication factor (i.e. the two approaches were applied to the same simulated samples), the other design factors were manipulated between replications.

Evaluation criteria. Factor structures were pooled using the indirect method as described above. The accuracy of the pooled factor structures were examined based on 1,000 replications using the same indices as the previous study: coefficient of alienation (*K*), (relative) bias (*B*), and mean squared error (*MSE*).

Results

The effects of the experimental manipulations on the three fit indices were analyzed using 3 (percentage of underextraction: 0%, 10%, 32%) x 2 (handling of underextraction: inclusion, exclusion) x 3 (reporting omissions: 0%, 10%, 25%) analyses of variance. The analysis of *K* was again based on the Fisher's *Z* transformed coefficients. The results of these analyses are summarized in Table 4.

Overall, the experimental manipulations hardly affected *K*, but had more pronounced effects on *MSE* (see Figure 5). Parameter accuracy was most strongly affected by the percentage of underextraction included in the meta-analyses (see hypothesis 5), η^2_G = .14 for *K* and η^2_G = .49 for *MSE*, and was larger when samples with underextraction were included (instead of discarded; see hypothesis 6), η^2_G = .14 and η^2_G = .44. Moreover, the interactions, η^2_G = .18 for *K* and η^2_G = .49 for *MSE*, indicated that parameter accuracy decreased strongest for meta-analyses that included a large proportion of samples where underextraction occurred (see Figure 5). In contrast, reporting omissions had a minor impact.

Separate analyses conducted for the salient and non-salient factor loadings (see Table S2 and Figure S2 in the online supplement) showed that reporting omissions (hypothesis 7) primarily affected the *MSE* of non-salient factor loadings (η_G^2 = .14) but hardly distorted salient loadings (η_G^2 = .00). In contrast, the inclusion of samples with underextraction (η_G^2 = .46) primarily biased salient factor loadings and to a lesser degree the accuracy of non-salient loadings (η_G^2 = .14). In terms of relative bias, these results translated into an underestimation of salient factor loadings of about 3% and 6% for meta-analyses including no underextraction or large underextraction, respectively. For non-salient loadings the respective values fell at 36% for both conditions. Thus, similar to the previous simulation the meta-analyses were unable to adequately estimate the magnitude of non-the salient factor loadings.

Discussion

The presented Monte Carlo simulation evaluated two new approaches for the meta-analysis of factor structures: the direct method subjected each loading individually to a meta-analysis, whereas the indirect method derived aggregated loadings from the pooled correlations reproduced by the sample factor loadings matrix. In general, both methods led to rather similar results: Meta-analyses of factor structures are most strongly affected by characteristics of the population factor structure; that is, the accuracy of the pooled factor loadings are predominately affected by the degree of model error and the communalities of the items. To a somewhat lesser degree respective results also depend on the sample size of the primary studies. Small-sample studies result in more biased factor loadings. However, the respective distortion in an individual study can be compensated in meta-analyses by including a large number of studies. With regard to the two approaches for aggregating factor structures the simulation identified small differences. Overall, compared to the direct method the indirect method was slightly superior and resulted in larger factor congruence and less parameter bias, particularly for meta-analyses of small-sample studies. Thus, in practice the choice of the meta-analytical method would be likely to be of minor consequence. However, the indirect method has several conceptual advantages (see also Table 2):

(a) It allows for the integration of primary studies with different number of extracted factors which is not possible in the direct method. (b) A second advantage is the flexibility of analyzing the pooled correlation matrix $\overline{\mathbf{R}^*}$ with different EFA models or CFA models. Thus, competing models can be tested and different methods (e.g., with respect to rotation) can be applied. (c) It is also possible to combine the indirect method with Becker's (1996) approach: Whenever the item correlation matrix **R** is presented in the primary study it can be used for the meta-analytical integration (as in MASEM). If only the sample loadings matrix is available the correlation matrix is reproduced using the indirect approach. The meta-analytical integration then combines the **R** and $\mathbf{R}^*$ matrices from the primary studies as described. It is also possible to code the type of correlation matrix (**R** versus $\mathbf{R}^*$) and analyze the type as a moderator variable. (d) Finally, moderator analyses could also be performed for other hypotheses (e.g., on cross-cultural measurement invariance) within the well-established framework of multi-group invariance testing (Davidov et al., 2014).

Directions for Future Research

The presented Monte Carlo simulations provided initial evidence for the suitability of two methods for the meta-analysis of factor structures. However, these results should be corroborated and extended by additional studies that could follow along two lines of research:

First, future research could refine the proposed meta-analytic methods. For example, the present study assumed homogenous population effects and, thus, presented methods for fixed-effects meta-analyses. However, the assumption that the population effect sizes are identical in all studies is rarely tenable in applied practice (Geyskens et al., 2009; Schmidt et al., 2009). Therefore, it is recommended to (a) study the accuracy of the indirect and direct approach for the pooling of heterogeneous factor loading matrices and (b) extend the introduced meta-analytic techniques to random effects models. Moreover, future research could also examine whether Fisher's *Z* transformations might improve the accuracy of the pooled factor loadings (cf. Hafdahl, 2009; 2010).

Second, future research should investigate further characteristics of the pooled primary studies that might affect the meta-analytic results. For example, the present study relied on maximum likelihood EFA. However, in practice researchers select various methods such as principal axis factoring or principal component analyses (Henson and Roberts, 2006). Simulation studies indicate that under certain conditions the choice of the estimator can have pronounced effects on factor pattern recovery (e.g., Briggs and MacCallum 2003; Velicier and Fava, 1998; de Winter and Dodou, 2012). Similarly, researchers can select among a variety of rotation criteria (cf. Browne, 2001) that have been shown to differentially affect resulting factor loading patterns (Schmitt and Sass, 2011). Therefore, future research on the direct approach is encouraged to study the effect of pooling factor loading matrices derived with different estimators using different rotation criteria.

Conclusions

The proposed new methods for the meta-analytical integration of factor structures offer alternatives to the Shafer (2005, 2006) approach and MASEM (Cheung, 2014, 2015) which are usually applied for this purpose. One advantage of the new methods is that they convey metric information about the magnitude of each loading. Taking into account the greater flexibility of use and the slightly better performance in the simulations especially the indirect method appears to be a promising method that deserves further attention.

References

Bakeman, R 2005, ‘Recommended effect size statistics for repeated measures designs’, *Behavior Research Methods*, vol. 37, no. 3, pp. 379-384. DOI: 10.3758/BF03192707

Becker, BJ 1992, ‘Using results from replicated studies to estimate linear models’, *Journal of Educational Statistics*, vol. 17, no. 4, pp. 341-362. DOI: 10.3102/10769986017004341

Becker, G 1996, ‘The meta-analysis of factor analyses: An illustration based on the cumulation of correlation matrices’, *Psychological Methods*, vol. 1, no. 4, pp. 341-353. DOI: 10.1037/1082-989X.1.4.341

Borg, I & Leutner, D 1985, ‘Measuring the similarity of MDS configurations’, *Multivariate Behavioral Research*, vol. 20, no. 3, pp. 325-334. DOI: 10.1207/s15327906mbr2003_6

Brannick, MT, Yang, LQ & Cafri, G 2011, ‘Comparison of weights for meta-analysis of *r* and *d* under realistic conditions’, *Organizational Research Methods*, vol. 14, no. 4, pp. 587-607. DOI: 10.1177/1094428110368725

Briggs, NE & MacCallum, RC 2003, ‘Recovery of weak common factors by maximum likelihood and ordinary least squares estimation’, *Multivariate Behavioral Research*, vol. 38, no. 1, pp. 25-56. DOI: 10.1207/S15327906MBR3801_2

Browne, MW 1972, ‘Orthogonal rotation to a partially specified target’, *British Journal of Mathematical and Statistical Psychology*, vol. 25, no. 1, pp. 115-120. DOI: 10.1111/j.2044-8317.1972.tb00482.x

Browne, MW 2001, ‘An overview of analytic rotation in exploratory factor analysis’, *Multivariate Behavioral Research*, vol. 36, no. 1, pp. 111-150. DOI: 10.1207/S15327906MBR3601_05

Bushman, BJ, Cooper, HM, & Lemke, KM 1991, ‘Meta-analysis of factor analyses: An illustration using the Buss-Durkee hostility inventory’, *Personality and Social Psychology Bulletin*, vol. 17, no. 3, pp. 344-349. DOI: 10.1177/0146167291173015

Cheung, MWL 2013, ‘Multivariate meta-analysis as structural equation models’, *Structural Equation Modeling*, vol. 20, no. 3, pp. 429-454. DOI: 10.1080/10705511.2013.797827

Cheung, MWL 2014 ‘Fixed-and random-effects meta-analytic structural equation modeling: Examples and analyses in R’, *Behavior Research Methods*, vol. 46, no. 1, pp. 29-40. DOI: 10.3758/s13428-013-0361-y

Cheung, MWL 2015, *Meta-analysis: A structural equation modeling approach*, 1st edn. Wiley, Chisester.

Cheung, MWL & Chan, W 2005, ‘Meta-analytic structural equation modeling: A two-stage approach’, *Psychological Methods,* vol. 10, pp. 40-64. DOI: 10.1037/1082-989X.10.1.40

Cheung, MWL. & Chan, W 2009, ‘A two-stage approach to fynthesizing covariance matrices in meta-analytic structural equation modeling’, *Structural Equation Modeling*, vol. 16, no. 1, pp. 28-53. DOI: 10.1080/10705510802561295

Davidov, E, Meuleman, B, Cieciuch, J, Schmidt, P & Billiet, J 2014, ‘Measurement equivalence in cross-national research’, *Annual Review of Sociology*, vol. 40, pp. 55-75. DOI: 10.1146/annurev-soc-071913-043137

Dunlap, WP, Cortina, JM, Vaslow, JB, & Burke, MJ 1996, ‘Meta-analysis of experiments with matched groups or repeated measures designs’, *Psychological Methods*, vol 1, no 2, pp. 170-177. DOI: 10.1037/1082-989X.1.2.170

Fabrigar, LR, Wegener, DT, MacCallum, RC, & Strahan, EJ 1999, ‘Evaluating the use of exploratory factor analysis in psychological research’, *Psychological Methods*, vol. 4, no. 3, pp. 272-299. DOI: 10.1037/1082-989X.4.3.272

Fava, JL, & Velicer, WF 1996, ‘The effects of underextraction in factor and component analyses’, *Educational and Psychological Measurement*, vol. 56, no.6, pp. 907-929. DOI: 10.1177/0013164496056006001

Ferguson, CJ 2009, ‘An effect size primer: A guide for clinicians and researchers’, *Professional Psychology: Research and Practice*, vol. 40, no. 5, pp. 532-538. DOI: 10.1037/a0015808

Field, AP 2001, ‘Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed-and random-effects methods’, *Psychological Methods*, vol. 6, no. 2, pp. 161-180. DOI: 10.1037/1082-989X.6.2.161

Field, AP 2005, ‘Is the meta-analysis of correlation coefficients accurate when population correlations vary?’, *Psychological Methods*, vol. 10, no. 4, pp. 444-467. DOI: 10.1037/1082-989X.10.4.444

Forero, CG, Maydeu-Olivares, A & Gallardo-Pujol, D 2009, ‘Factor analysis with ordinal indicators: A Monte-Carlo study comparing DWLS and ULS estimation’ *Structural Equation Modeling*, vol. 16, no. 4, pp. 625-641. DOI: 10.1080/10705510903203573

Furlow, CF & Beretvas, SN 2005, ‘Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data’, *Psychological Methods*, vol. 10, no. 2, pp. 227-254. DOI: 10.1037/1082-989X.10.2.227

Gebhardt, F 1968, ‘The similarity of factor matrices’, *Psychologische Beiträge*, vol. 10, no., 4, pp. 591-599.

Geyskens, I, Krishnan, R, Steenkamp, JBEM & Cunha, PV 2009, ‘A review and evaluation of meta-analysis practices in management research’ *Journal of Management*, vol. 35, no. 2, pp. 393-419. DOI: 10.1177/0149206308328501

Gnambs, T 2013, ‘The elusive general factor of personality: The acquaintance effect’, *European Journal of Personality,* vol. 27, no. 5, pp. 507-520. DOI: 10.1002/per.1933

Gnambs, T 2014, ‘A meta-analysis of dependability coefficients (test-retest reliabilities) for measures of the Big Five’, *Journal of Research in Personality,* vol. 52, pp. 20-28. DOI: 10.1016/j.jrp.2014.06.003

Gnambs, T 2015, ‘Facets of measurement error for scores of the Big Five: Three reliability generalizations’, *Personality and Individual Differences*, vol. 84, pp. 84-89. DOI: 10.1016/j.paid.2014.08.019

Guadagnoli, E & Velicer, WF 1988, ‘Relation of sample size to the stability of component patterns’, *Psychological Bulletin*, vol. 103, no. 2, pp. 265-275. DOI: 10.1037/0033-2909.103.2.265

Hafdahl, AR 2007, ‘Combining correlation matrices: Simulation analysis of improved fixed-effects methods’, *Journal of Educational and Behavioral Statistics*, vol. 32, no. 2, pp. 180-205. DOI: 10.3102/1076998606298041

Hafdahl, AR 2009, ‘Improved Fisher z estimators for univariate random-effects meta-analysis of correlations’, *British Journal of Mathematical and Statistical Psychology*, vol. 62, no. 2, pp. 233-261. DOI: 10.1348/000711008X281633

Hafdahl, AR 2010, ‘Random-effects meta-analysis of correlations: Monte Carlo evaluation of mean estimators’, *British Journal of Mathematical and Statistical Psychology*, vol. 63, no. 1, pp. 227-254. DOI: 10.1348/000711009X431914

Hafdahl, AR & Williams, MA 2009, ‘Meta-analysis of correlations revisited: Attempted replication and extension of Field’s (2001) simulation studies’, *Psychological Methods*, vol. 14, no. 1, pp. 24-42. DOI: 10.1037/a0014697

Hedges, LV 1981, ‘Distribution theory for Glass's estimator of effect size and related estimators’, *Journal of Educational Statistics*, vol. 6, no. 2, 107-128. DOI: 10.3102/10769986006002107

Hedges, LV 1983, 'A random effects model for effect sizes', *Psychological Bulletin,* vol. 93, no. 2, pp. 388-395. DOI: 10.1037/0033-2909.93.2.388

Hedges, LV & Vevea, JL 1998, ‘Fixed- and random-effects models in meta-analysis’, *Psychological Methods*, vol. 3, no. 4, pp. 486-504. DOI: 10.1037/1082-989X.3.4.486

Henson, RK & Roberts, JK 2006, ‘Use of exploratory factor analysis in published research’, *Educational and Psychological Measurement*, vol. 66, no. 3, pp. 393-416. DOI: 10.1177/0013164405282485

Hogarty, KY, Hines, CV, Kromrey, JD, Ferron, JM & Mumford, KR 2005, ‘The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination’, *Educational and Psychological Measurement*, vol. 65, no. 2, pp. 202-226. DOI: 10.1177/0013164404267287

Hong, S 1999, ‘Generating correlation matrices with model error for simulation studies in factor analysis: A combination of the Tucker-Koopman-Linn model and Wijsman's algorithm’, *Behavior Research Methods, Instruments, & Computers*, vol. 31, no. 4, pp. 727-730. DOI: 10.3758/BF03200754

Hoogland, JJ & Boomsma, A 1998, ‘Robustness studies in covariance structure modeling: An overview and a meta-analysis’, *Sociological Methods & Research*, vol. 26, no. 3, pp. 329-367. DOI: 10.1177/0049124198026003003

Huang, C & Dong, N 2012, ‘Factor structures of the Rosenberg self-esteem scale: A meta-analysis of pattern matrices’, *European Journal of Psychological Assessment,* vol. 28, no. 2, pp. 132-138. DOI: 10.1027/1015-5759/a000101

Kaiser, HF, Hunka, S & Bianchini, JC 1971, ‘Relating factors between studies based upon different individuals’, *Multivariate Behavioral Research*, vol. 6, no. 4, pp. 409-442. DOI: 10.1207/s15327906mbr0604_3

Kasten, N & Freund, PA 2015, ‘A meta-analytical multilevel reliability generalization of Situational Judgment Tests (SJTs)’, *European Journal of Psychological Assessment*. Advance online publication. DOI: 10.1027/1015-5759/a000250

MacCallum, RC & Tucker, LR 1991, ‘Representing source of error in the common-factor model: Implications for theory and practice’ *Psychological Bulletin*, vol. 109, no. 3, pp. 502-511. DOI: 10.1037/0033-2909.109.3.502

MacCallum, RC, Widaman, KF, Preacher, KJ & Hong, S 2001, 'Sample size in factor analysis: The role of model error', *Multivariate Behavioral Research,* vol. 36, no. 4, pp. 611-637. DOI: 10.1207/S15327906MBR3604_06

MacCallum, RC, Widaman, KF, Zhang, S & Hong, S 1999, 'Sample size in factor analyses', *Psychological Methods*, vol. 4, no. 1, pp. 84-99. DOI: 10.1037/1082-989X.4.1.84

McDonald, RP 1999, *Test theory: A unified treatment*, 1st edn. Lawrence Erlbaum, Mahwah.

Meredith, W 1993, 'Measurement invariance, factor analysis and factorial invariance', *Psychometrika*, vol. 58, no. 4, pp. 525-543. DOI: 10.1007/BF02294825

Mood, AM, Graybill, FA & Boes, DC 1974, *Introduction to the theory of statistics*, 3rd edn. McGraw-Hill, New York.

Myers, ND, Jin, Y, Ahn, S, Celimli, S & Zopluoglu, C 2015, 'Rotation to a partially specified target matrix in exploratory factor analysis in practice' *Behavior Research Methods*, vol. 47, no. 1, pp. 494-505. DOI: 10.3758/s13428-014-0486-7

Olejnik, S & Algina, J 2003, 'Generalized eta and omega squared statistics: Measures of effect size for some common research designs', *Psychological Methods*, vol. 8, no. 4, pp. 434-447. DOI: 10.1037/1082-989X.8.4.434

Olkin, I. & Finn, J 1995, 'Correlations redux', *Psychological Bulletin*, vol. 118, no. 1, pp. 155-164. DOI: 10.1037/0033-2909.118.1.155

Preacher, KJ & MacCallum, RC 2002, 'Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes', *Behavior Genetics*, vol. 32, no. 2, pp. 153-161. DOI: 10.1023/A:1015210025234

Prevost, AT, Mason, D, Griffin, S, Kinmonth, AL, Sutton, S & Spiegelhalter, D 2007, ‘Allowing for correlations between correlations in random-effects meta-analysis of correlation matrices’, *Psychological Methods*, vol. 12, no. 4, pp. 434-450. DOI: 10.1037/1082-989X.12.4.434

Rizzo, ML 2008, *Statistical computing in R.* 1st edn. Chapman & Hall, Boca Raton.

Salgado, JF & Táuriz, G 2014, ‘The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies’, *European Journal of Work and Organizational Psychology*, vol. 23, no. 1, pp. 3-30. DOI: 10.1080/1359432X.2012.716198

Schermelleh-Engel, K, Moosbrugger, H & Müller, H 2003, ‘Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures’, *Methods of Psychological Research Online*, vol. 8, no 2, pp. 23-74.

Schmidt, FL, Oh, IS & Hayes, TL 2009, ‘Fixed-versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results’, *British Journal of Mathematical and Statistical Psychology*, vol. 62, no. 1, pp. 97-128. DOI: 10.1348/000711007X255327

Schmitt, TA & Sass, DA 2011, ‘Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations’, *Educational and Psychological Measurement*, vol. 71, no. 1, pp. 95-113. DOI: 10.1177/0013164410387348

Schönemann, PH 1966 ‘A generalized solution of the orthogonal Procrustes problem’, *Psychometrika*, vol. 31, no 1., pp. 1-10. DOI: 10.1007/BF02289451

Shafer, AB 2005, ‘Meta-analysis of the Brief Psychiatric Rating Scale factor structure’, *Psychological Assessment,* vol. 17, no. 3, pp. 324-335. DOI: 10.1037/1040-3590.17.3.324

Shafer, AB 2006, ‘Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung’, *Journal of Clinical Psychology,* vol. 62, no. 1, pp. 123-146. DOI: 10.1002/jclp.20213

Tucker, LR 1951, *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Department of the Army, Washington.

Tucker, LR, Koopman, RF,& Linn, RL 1969, ‘Evaluation of factor analytic research procedures by means of simulated correlation matrices’, *Psychometrika*, vol. 34, no. 4, pp. 421-459. DOI: 10.1007/BF02290601

Van Iddekinge, CH, Roth, PL, Raymark, PH & Odle-Dusseau, HN 2012, ‘The criterion-related validity of integrity tests: An updated meta-analysis’. *Journal of Applied Psychology*, vol. 97, no. 3, pp. 499-530. DOI: 10.1037/a0021196

Velicer, WF & Fava, JL 1998, ‘Effects of variable and subject sampling on factor pattern recovery’, *Psychological Methods*, vol. 3, no. 2, pp. 231-251. DOI: 10.1037/1082-989X.3.2.231

Viechtbauer, W, López-López, JA, Sánchez-Meca, J & Marín-Martínez, F 2014, ‘A comparison of procedures to test for moderators in mixed-effects meta-regression models’, *Psychological Methods*. Advance online publication. DOI: 10.1037/met0000023 de Winter, JCF, Dodou, D & Wieringa, PA 2009, ‘Exploratory factor analysis with small sample sizes’, *Multivariate Behavioral Research*, vol. 44, no. 2, pp. 147-181. DOI: 10.1080/00273170902794206

de Winter, JCF & Dodou, D 2012, ‘Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size’, *Journal of Applied Statistics*, vol. 39, no. 4, pp. 695-710. DOI: 10.1080/02664763.2011.610445

Wood, JM, Tataryn, DJ & Gorsuch, RL 1996, ‘Effects of under-and overextraction on principal axis factor analysis with varimax rotation’, *Psychological Methods*, vol. 1, no. 4, pp. 354-365. DOI: 10.1037/1082-989X.1.4.354

Worley, JA, Vassar, M, Wheeler, DL & Barnes, LL 2008, ‘Factor structure of scores from the Maslach Burnout Inventory: A review and meta-analysis of 45 exploratory and confirmatory factor-analytic studies’, *Educational and Psychological Measurement,* vol. 68, no. 5, pp. 797-823. DOI: 10.1177/0013164408315268

Table 1.

Overview of Simulation Procedure by Meta-Analytic Approach

	Direct method	*Indirect method*	*Control condition*
Step 1:	Six $p \times p$ population correlation matrices **P** (see Appendix B) are constructed which *approximate* a factor structure $\mathbf{\Lambda}$ with q correlated major common factors and a $q \times q$ matrix of correlations between major factors $\mathbf{\Phi}$. Here $p = 10$, $q = 2$, and $\varphi_{12} = .30$ are invariant. Five items have salient loadings on factor 1 and five items on factor 2. The idealized $p \times q$ pattern target matrix **T** contains five rows (1, 0) and five rows (0, 1). The communalities (low, wide, or high) and the model error (small or high) are manipulated.		
Step 2:	Data for a meta-analysis are generated by drawing k random samples of size n from a multivariate normal distribution with the correlation structure **P.** This yields k correlation matrices **R**. Here k and n are manipulated.		
Step 3:	An exploratory maximum likelihood factor analysis with two oblimin rotated factors is performed on each **R** resulting in a $p \times q$ pattern matrix **L** and a $q \times q$ matrix of correlations between factors **F** for each sample.		
Step 4:	(a) The pattern matrices **L** are rotated to the target matrix **T** using an orthogonal Procrustes rotation and yield matrices $\tilde{\mathbf{L}}$.	(a) From the pattern matrices **L** the sample correlation matrices $\mathbf{R}^*$ are reproduced via $\mathbf{R}^* = \mathbf{LFL}'$.	(a) The k sample correlation matrices **R** are meta-analytically aggregated into $\overline{\mathbf{R}}$.
	(b) The k pattern matrices $\tilde{\mathbf{L}}$ are meta-analytically aggregated into $\overline{\mathbf{L}}$.	(b) The k matrices of the reproduced correlations $\mathbf{R}^*$ are meta-analytically aggregated into $\overline{\mathbf{R}^*}$.	(b) $\overline{\mathbf{R}}$ is subjected to an exploratory maximum likelihood factor analysis with two oblimin rotated factors resulting in pattern matrix $\overline{\mathbf{L}}$.
		(c) $\overline{\mathbf{R}^*}$ is subjected to an exploratory maximum likelihood factor analysis with two oblimin rotated factors resulting in pattern matrix $\overline{\mathbf{L}}$.	
Step 5:	The pattern matrix $\overline{\mathbf{L}}$ is rotated to the population target matrix $\mathbf{\Lambda}$ by means of an orthogonal Procrustes rotation yielding matrix $\hat{\mathbf{\Lambda}}$.		
Step 6:	Steps 2 to 5 are repeated 1,000 times for each experimental condition. Fit indices between the two matrices $\hat{\mathbf{\Lambda}}$ and $\mathbf{\Lambda}$ (coefficient of alienation, *K*; root mean squared error, *RMSE*; and bias, *B*) are computed.		

Table 2.

Comparison of Meta-Analytic Approaches

	Direct method	Indirect method	MASEM
What are the required effect sizes from the primary studies?	Factor loadings	Factor loadings [d]	Item-level correlations
Can loading matrices with a different number of factors be pooled?	no	yes	-
Can competing configural models be examined?	no	yes	yes
Are estimates of pooled factor loadings available?	yes	yes	yes
How can moderator hypotheses be addressed?	Meta-regression analyses [a] for each factor loading individually	Multi-group analyses [b], gradients of factor loadings [c]	Multi-group analyses [b], gradients of factor loadings [c]

Note. [a] see Viechtbauer et al. (2014); [b] see Davidov et al. (2014); [c] see Gnambs (2013); [d] Although not examined in these simulations it is in principle possible to meta-analytically combine the correlations matrices reproduced form the loadings with item-level correlation matrices as in MASEM.

Table 3.

ANOVA Results for the Accuracy of Aggregated Factor Loading Matrices in Study I

	Coefficient of alienation K			*Bias B*			*Relative Bias* B_r			*Mean squared error MSE*		
	F	*df*	η^2_G	*F*	*df*	η^2_G	*F*	*df*	η^2_G	*F*	*df*	η^2_G
Aggregation method (*m*)	42,079	2	.068	121,230	2	.101	118,056	2	.176	27,129	2	.049
Communality (*h*)	289,660	2	.907	18,905	2	.400	6,614	2	.181	35,136	2	.539
Model error (*e*)	295,531	1	.832	8534	1	.015	843	1	.014	202,762	1	.771
Number of samples (*k*)	4,697	2	.136	2†	2	.000	9	2	.000	4,281	2	.125
Sample size (*n*)	8,597	2	.224	18	2	.001	805	2	.026	7,737	2	.205
m x *h*	14,041	4	.046	55,211	4	.093	59,283	4	.177	6,450	4	.024
m x *e*	300	2	.001	4,403	2	.004	3,287	2	.006	414	2	.001
m x *k*	86	4	.000	93	4	.000	85†	4	.000	7	4	.000
m x *n*	10,739	4	.036	24,698	4	.044	24,519	4	.082	8,205	4	.030
h x *e*	116,166	2	.796	66	2	.002	1,070	2	.035	71,093	2	.703
h x *k*	2,299	4	.080	5	4	.000	11	4	.001	953	4	.060
h x *n*	1,985	4	.118	293	4	.020	641	4	.041	1,1780	4	.073
e x *k*	38	2	.001	0†	2	.000	0†	2	.000	67	2	.002
e x *n*	220	2	.007	8	2	.000	18	2	.001	21	2	.001
k x *n*	1,207	4	.075	1†	4	.000	4†	4	.000	1,265	4	.078

Note. Results for higher-order interactions with $\eta^2_G < .04$ are not presented. All effects except those marked with †are significant at $p < .001$.

Table 4.

ANOVA Results for the Accuracy of Aggregated Factor Loading Matrices in Study II

	Coefficient of alienation K			*Bias B*			*Relative bias B_r*			*Mean squared error MSE*		
	F	*df*	η^2_G	*F*	*df*	η^2_G	*F*	*df*	η^2_G	*F*	*df*	η^2_G
Percentage of underextraction (*u*)	903	2	.139	11,579	2	.639	752	2	.140	7,215	2	.486
Handling of underextraction (*h*)	7,109	1	.135	38,050	1	.570	20,356	1	.053	16,919	1	.436
Reporting omissions (*o*)	1,169	2	.173	5,032	2	.435	15,168	2	.767	135	2	.017
u x *h*	5,009	2	.181	19,444	2	.575	10,839	2	.056	10,614	2	.492
u x *o*	34	4	.012	5	4	.001	$2^{\dagger}$	4	.001	9	4	.002
h x *o*	408	2	.018	14	2	.001	36	2	.000	40	2	.004
u x *h* x *o*	198	4	.017	7	4	.001	19	4	.000	23	4	.004

Note. All effects except those marked with †are significant at $p < .001$.

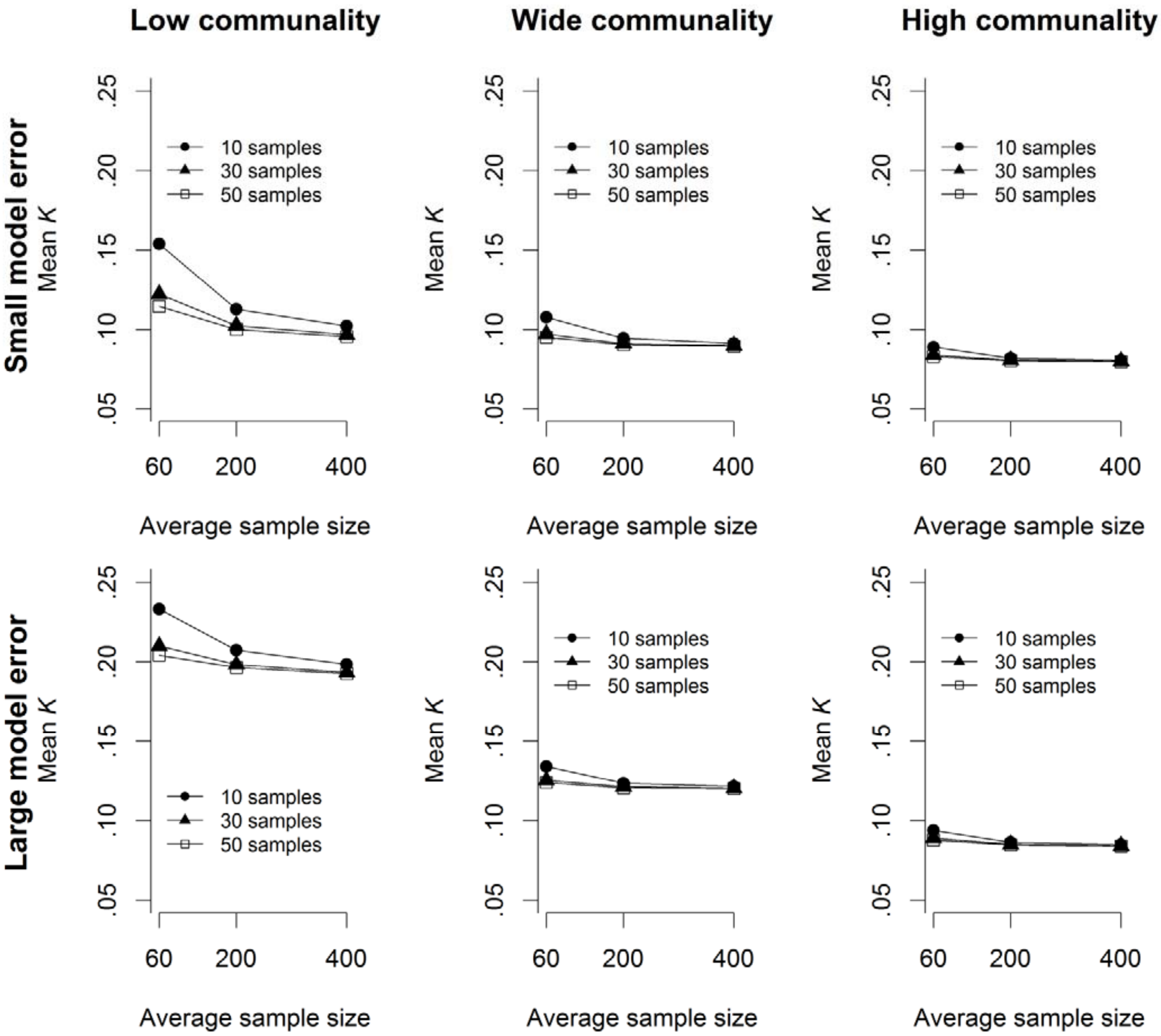

Figure 1. Mean coefficients of alienation *K* by experimental conditions in Study I

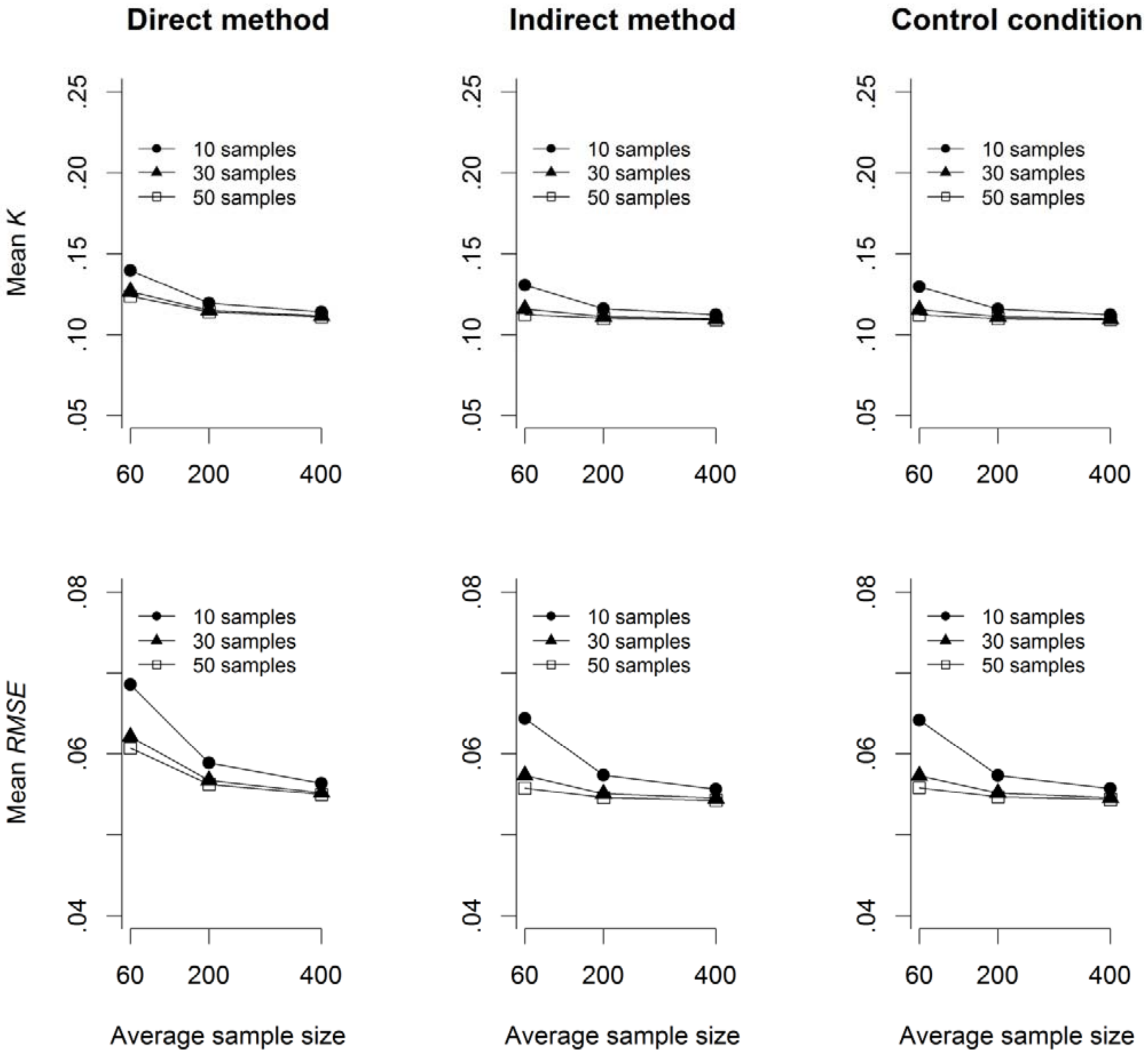

Figure 2. Mean coefficients of alienation *K* (top row) and root mean squared errors *RMSE* (bottom row) by meta-analytical method in Study I

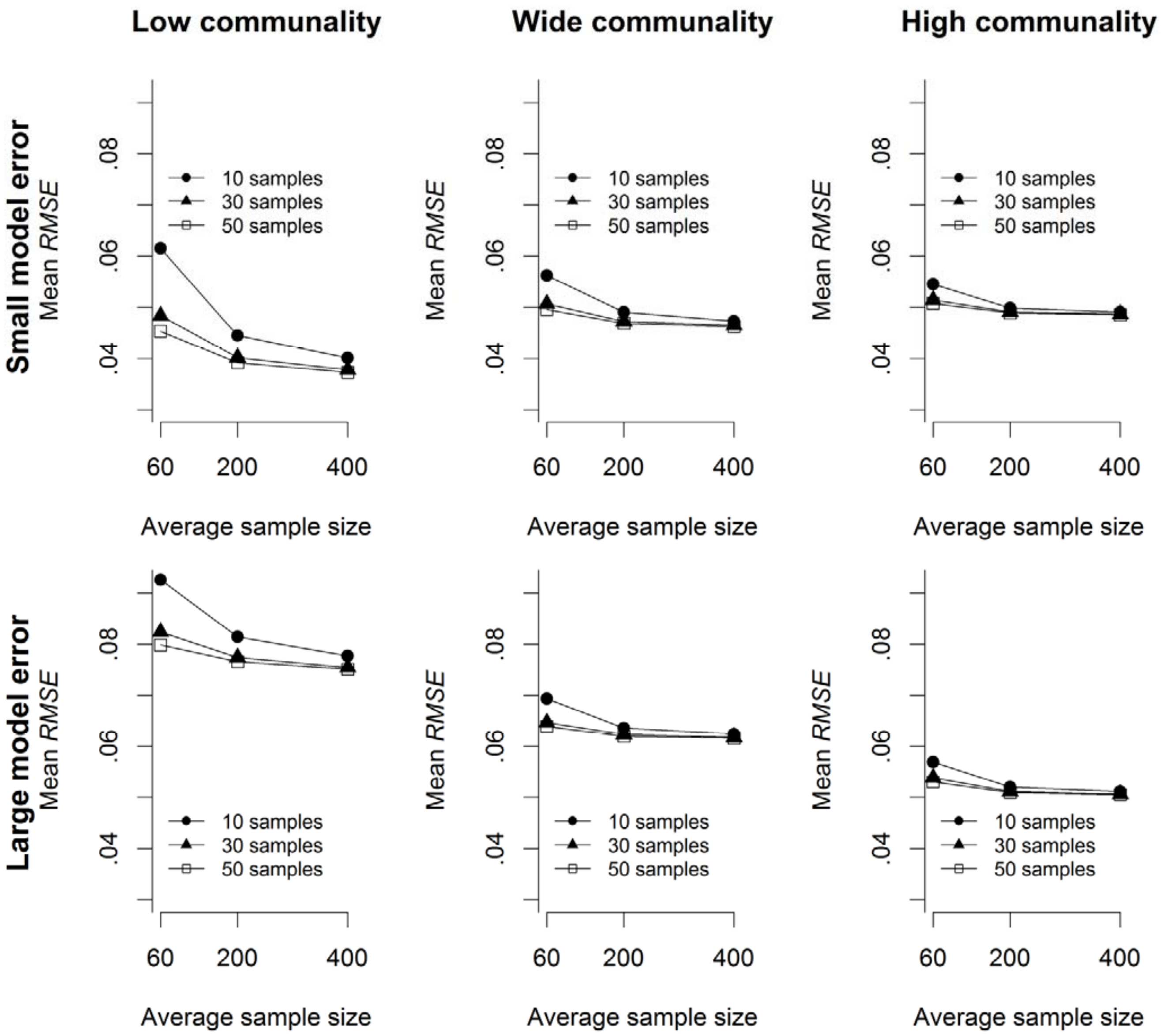

Figure 3. Mean root mean squared error *RMSE* by experimental conditions in Study I

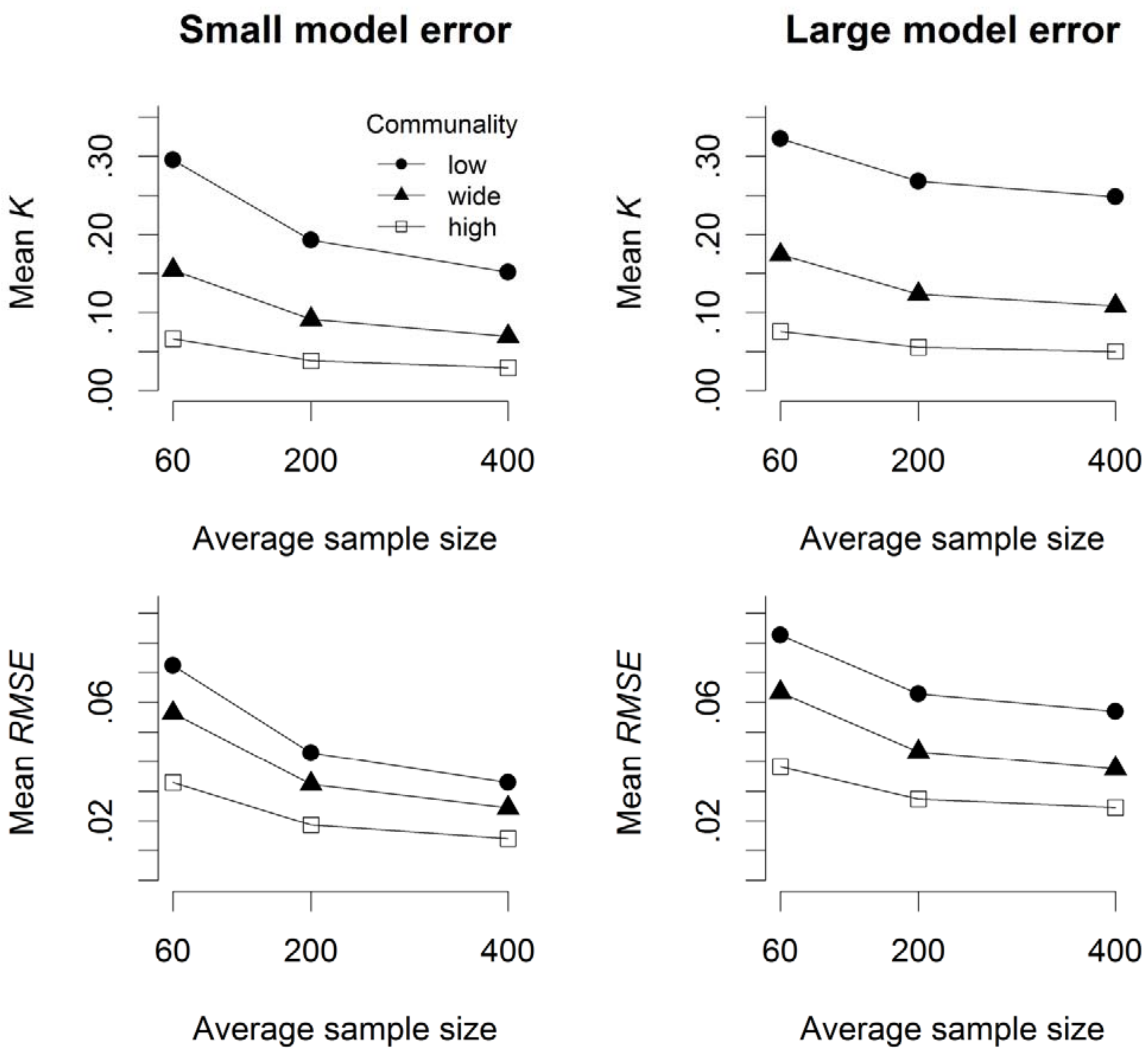

Figure 4. Mean coefficients of alienation *K* (top row) and root mean squared errors *RMSE* (bottom row) for the accuracy of the reproduced correlation matrices

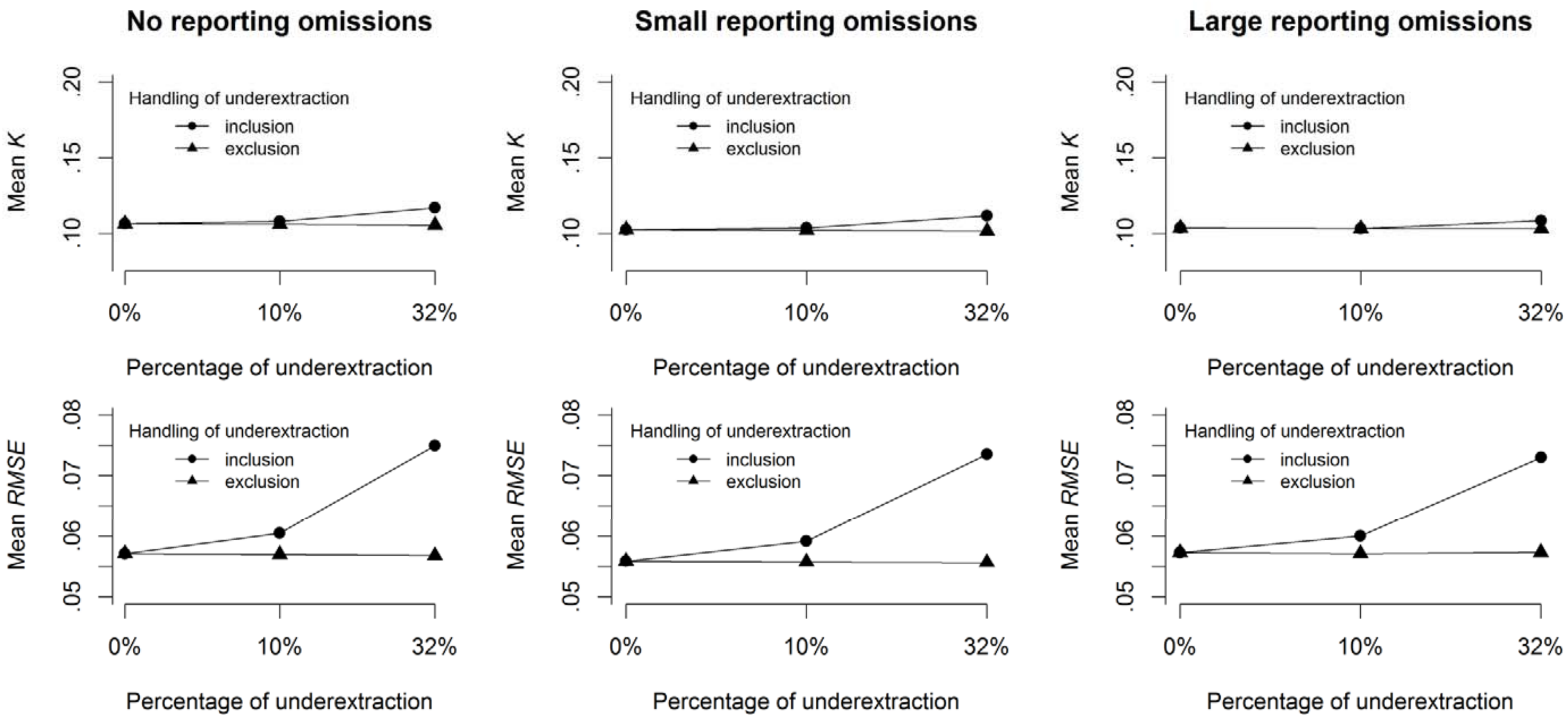

Figure 5. Mean parameter accuracy by experimental conditions in Study II

Appendix A

Major factor loadings for the population model in the three experimental conditions

	Low communality			Wide communality			High communality		
Item	F_1	F_2	h^2	F_1	F_2	h^2	F_1	F_2	h^2
I_1	.52	.17	.36	.52	.17	.36	.76	.25	.76
I_2	.39	.13	.20	.76	.25	.76	.68	.23	.60
I_3	.52	-.17	.25	.52	-.17	.25	.84	-.28	.65
I_4	.50	.17	.33	.50	.17	.33	.75	.25	.73
I_5	.55	-.18	.28	.86	-.29	.68	.86	-.29	.68
I_6	.18	.55	.39	.18	.53	.37	.26	.78	.79
I_7	-.19	.58	.31	-.29	.88	.71	-.29	.88	.71
I_8	-.16	.49	.22	-.16	.49	.22	-.27	.82	.62
I_9	-.18	.54	.27	-.27	.82	.62	-.29	.86	.67
I_{10}	.18	.53	.37	.26	.77	.77	.26	.77	.77

Appendix B

Population correlation matrices by experimental conditions

		I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	
Low communality, small model error	I_1		.33	.12	.37	.29	.29	.07	.02	.07	.33	Low communality, large model error
	I_2	.21		.06	.40	.35	.19	.17	.10	.12	.23	
	I_3	.26	.19		.17	.23	.08	-.14	-.13	-.07	.03	
	I_4	.37	.22	.24		.39	.28	.13	.08	.12	.20	
	I_5	.26	.21	.30	.24		.06	.00	-.07	.04	.06	
	I_6	.31	.18	.06	.29	.10		.20	.07	.27	.46	
	I_7	.12	.05	-.10	.12	-.10	.30		.36	.35	.31	
	I_8	.04	.08	-.07	.06	-.07	.21	.25		.23	.16	
	I_9	.12	.02	-.08	.10	-.10	.30	.34	.22		.25	
	I_{10}	.27	.24	.08	.27	.06	.36	.27	.22	.27		
Wide communality, small model error	I_1		.71	.56	.77	.54	.60	.16	.12	.19	.65	Wide communality, large model error
	I_2	.67		.51	.67	.45	.54	.15	.15	.17	.55	
	I_3	.57	.51		.58	.73	.16	-.24	-.28	-.28	.19	
	I_4	.74	.67	.56		.55	.59	.14	.11	.20	.60	
	I_5	.59	.54	.67	.57		.15	-.22	-.23	-.26	.14	
	I_6	.57	.51	.19	.57	.17		.62	.58	.60	.82	
	I_7	.18	.16	-.25	.18	-.26	.61		.68	.69	.62	
	I_8	.17	.15	-.21	.16	-.23	.59	.65		.60	.53	
	I_9	.15	.15	-.24	.19	-.24	.61	.69	.64		.59	
	I_{10}	.59	.52	.18	.57	.16	.78	.60	.55	.58		

High communality, small model error	I_1		.54	.22	.35	.40	.28	.14	.04	.13	.43	High communality, large model error
	I_2	.54		.35	.51	.59	.42	.21	.12	.18	.60	
	I_3	.22	.35		.23	.43	.07	-.14	-.08	-.15	.10	
	I_4	.35	.51	.23		.37	.24	.09	.05	.11	.39	
	I_5	.40	.59	.43	.37		.12	-.25	-.13	-.22	.17	
	I_6	.28	.42	.07	.24	.12		.43	.25	.38	.51	
	I_7	.14	.21	-.14	.09	-.25	.43		.43	.66.	.60	
	I_8	.04	.12	-.08	.05	-.13	.25	.42		.36	.31	
	I_9	.13	.18	-.15	.11	-.22	.38	.66	.36		.56	
	I_{10}	.43	.60	.10	.39	.17	.51	.60	.31	.56		

Online supplement for

“Parameter Accuracy in Meta-Analyses of Factor Structures”

Table S1.

ANOVA Results for the Accuracy of Salient and Non-Salient Factor Loadings in Study I

	Bias B						*Relative bias B_r*					
	Salient factor loadings			Non-salient factor loadings			Salient factor loadings			Non-salient factor loadings		
	F	*df*	η^2_G	*F*	*df*	η^2_G	*F*	*df*	η^2_G	*F*	*df*	η^2_G
Aggregation method (*m*)	65,156	2	.021	130,371	2	.193	55,296	2	.020	116,193	2	.183
Communality (*h*)	13,326	2	.327	81,555	2	.732	2,930	2	.096	8,639	2	.223
Model error (*e*)	81	1	.001	778	1	.013	2,547	1	.044	120	1	.002
Number of samples (*k*)	3	2	.000	12	2	.000	$2^\dagger$	2	.000	12	2	.000
Sample size (*n*)	1,063	2	.037	1,525	2	.049	745	2	.026	1,374	2	.044
m x *h*	33,163	4	.022	48,388	4	.151	33,697	4	.025	56,198	4	.179
m x *e*	17,058	2	.006	4,018	2	.007	17,424	2	.006	3,780	2	.007
m x *k*	12	4	.000	87	4	.000	10	4	.000	80	4	.000
m x *n*	19,978	4	.013	29,017	4	.096	16,655	4	.012	24,923	4	.088
h x *e*	2,565	2	.085	2,084	2	.065	3,893	2	.124	1,919	2	.060
h x *k*	$0^\dagger$	4	.000	10	4	.001	$0^\dagger$	4	.000	12	4	.001
h x *n*	9	4	.001	541	4	.035	22	4	.002	684	4	.044
e x *k*	$0^\dagger$	2	.000	$0^\dagger$	2	.000	$0^\dagger$	2	.000	$0^\dagger$	2	.000
e x *n*	19	2	.001	40	2	.001	27	2	.002	32	2	.001
k x *n*	83	4	.000	6	4	.000	2	4	.000	$2^\dagger$	6	.000

Note. Results for higher-order interactions with $\eta^2_G < .04$ are not presented. All effects except those marked with †are significant at $p < .001$.

Table S1 (continued).

	Mean square error MSE					
	Salient factor loadings			Non-salient factor loadings		
	F	*df*	η_G^2	*F*	*df*	η_G^2
Aggregation method (*m*)	161	2	.000	58,103	2	.095
Communality (*h*)	94,809	2	.764	4,298	2	.126
Model error (*e*)	247,498	1	.809	26,182	1	.305
Number of samples (*k*)	2,038	2	.065	2,605	2	.080
Sample size (*n*)	2,250	2	.071	6,736	2	.184
m x *h*	111	4	.000	14,285	4	.049
m x *e*	15	2	.000	938	2	.002
m x *k*	14	4	.000	21	4	.000
m x *n*	60	4	.000	17,243	4	.059
h x *e*	81,968	2	.737	12,327	2	.292
h x *k*	571	4	.038	460	4	.030
h x *n*	243	4	.016	1,234	4	.076
e x *k*	29	2	.001	44	2	.001
e x *n*	13	2	.000	85	2	.003
k x *n*	613	4	.040	758	4	.048

Table S2.

ANOVA Results for the Accuracy of Salient and Non-Salient Factor Loadings in Study II

	Bias B			*Relative bias* B_r			*Mean square error MSE*		
Salient loadings									
	F	*df*	η^2_G	*F*	*df*	η^2_G	*F*	*df*	η^2_G
Percentage of underextraction (*u*)	13,105	2	.645	11,207	2	.623	9,954	2	.541
Handling of underextraction (*h*)	37,656	1	.611	37,276	1	.582	16,209	1	.457
Reporting omissions (*o*)	69	2	.010	101	2	.015	21	2	.002
u x *h*	19,175	2	.616	19,030	2	.587	10,491	2	.522
u x *o*	5	4	.001	6	4	.002	6	4	.001
h x *o*	22	2	.002	27	2	.002	10	2	.001
u x *h* x *o*	10	4	.002	13	4	.002	7	4	.001
Non-salient loadings									
	F	*df*	η^2_G	*F*	*df*	η^2_G	*F*	*df*	η^2_G
Percentage of underextraction (*u*)	64	2	.014	47	2	.010	377	2	.071
Handling of underextraction (*h*)	21,946	1	.064	19,477	1	.046	15,523	1	.140
Reporting omissions (*o*)	14,372	2	.757	14,145	2	.754	825	2	.143
u x *h*	10,864	2	.063	9,520	2	.045	8,165	2	.147
u x *o*	$0^\dagger$	4	.000	$0^\dagger$	4	.000	10	4	.004
h x *o*	83	2	.001	$2^\dagger$	2	.000	544	2	.011
u x *h* x *o*	38	4	.000	$1^\dagger$	4	.000	272	4	.011

Note. All effects except those marked with †are significant at $p < .001$.

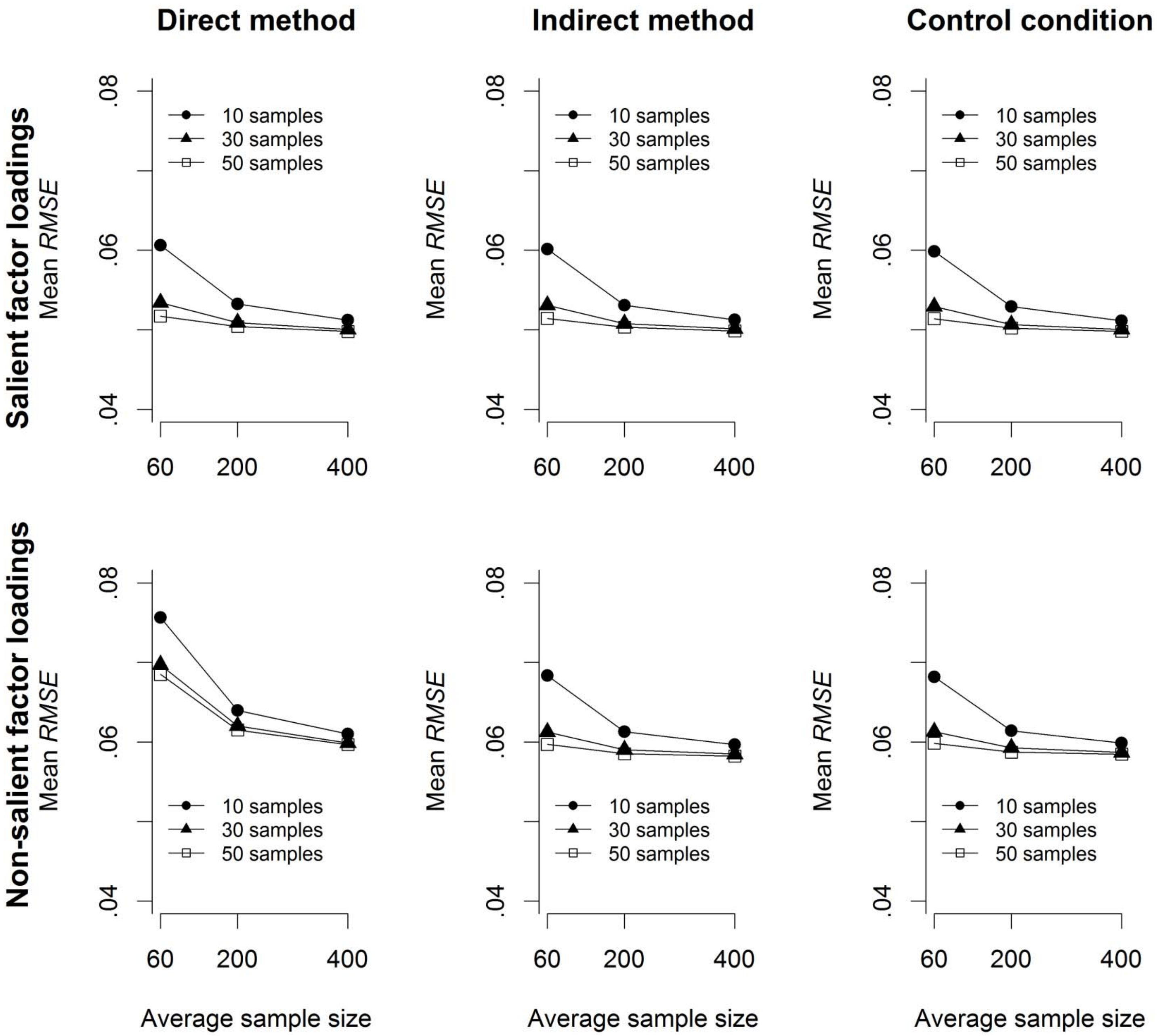

Figure S1. Mean root mean square error *RMSE* for salient (bottom row) and non-salient (top row) factor loadings by meta-analytical method in Study I

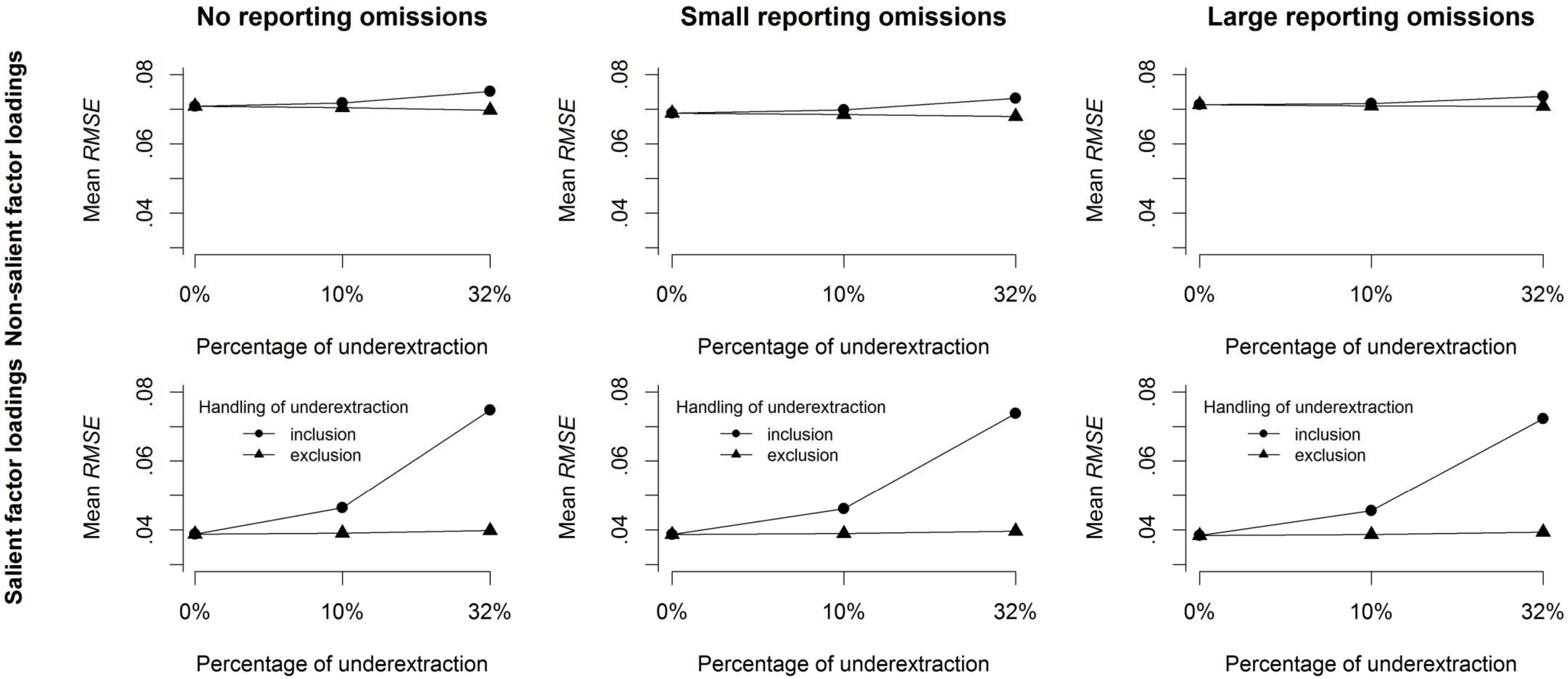

Figure S2. Mean root mean square error *RMSE* for salient (bottom row) and non-salient (top row) factor loadings by experimental conditions in Study II