Chapter 27

Exploratory Structural Equation Modeling

Alexandre J. S. Morin

Substantive-Methodological Synergy Research Laboratory, Department of Psychology, Concordia University, Canada

Acknowledgment

The author warmly thanks István Tóth-Király for his help with proof-reading, and for his invaluable contribution to the functioning of the Substantive-Methodological Synergy Research Laboratory, without which I would never have been able to find time to prepare this chapter

This is the prepublication version of the following chapter:

Morin, A.J.S. (In Press). Exploratory structural equation modeling. In R.H. Hoyle (Ed)., *Handbook of Structural Equation Modeling, Second Edition*. Guilford.

Exploratory Factor Analysis (EFA) was invented over a century ago (Spearman, 1904) to investigate the structure of psychological constructs: Unobservable entities (such as intelligence or self-esteem) inferred from a series of observed indicators (often questionnaire items). EFA was (and still is) designed to represent these psychological constructs through latent factors assumed to “cause” the covariance among a set of observed indicators considered to reflect the latent construct (Bollen & Hoyle, Chapter 5, this volume). Although early applications of EFA sought to statistically explore the structure of psychological constructs for the first time, EFA was not initially described as “exploratory” but only referred to as “factor analysis.” The “exploratory” label came later, following the development of Confirmatory Factor Analysis (CFA) and Structural Equation Modeling (SEM).

The “confirmatory” label associated with these methods came from their ability to rely on an *a priori* specification of the structure of the latent factors and to verify the extent to which this specification matched the data (Brown, Chapter 14, this volume). By contrast, the “exploratory” label came to be associated with EFA, to reflect the fact that EFA factors are defined by all possible indicators through rotation procedures and interpreted via the examination of the relative strength of factor-indicators associations, classified as loadings (associations between factors and their main indicators) or cross-loadings (the remaining, weaker, associations). With CFA/SEM came major statistical developments unrelated to this distinction, allowing researchers to estimate chains of relations between latent factors corrected for unreliability, longitudinal trajectories and change, measurement equivalence (i.e., invariance) across subpopulations or occasions, model fit assessment procedures, and so on. Because these developments were limited to CFA/SEM, far more than because of the free estimation of cross-loadings, EFA progressively became a second-class citizen seen as mainly useful for preliminary analyses.

Accumulating evidence has recently revealed that, by forcing all cross-loadings to be exactly zero, CFA/SEM carries its own load of problems. Marsh et al. (2005) noted that it was virtually impossible to obtain an acceptable level of fit for complex measures including “multiple factors (e.g., 5-10), each measured with a reasonable number of items (e.g., at least 5-10 per scale) so that there are at least 50 items overall” (p. 325). Others noted that many well-established measures presented a well-defined EFA structure that was almost impossible to replicate using CFA (Marsh et al., 2009; McCrae et al., 1996). Statistical research has also revealed that forcing even substantively unimportant cross-loadings (i.e., as small as .100) to be exactly zero resulted in biased estimates of the latent constructs themselves, leading to inflated factor correlations (Asparouhov et al., 2015) and biased estimates of factor regressions (Mai et al., 2018). These studies also revealed that freely estimating cross-loadings, even when none were necessary, still resulted in accurate estimates of the latent factors (Asparouhov et al., 2015).

The development of exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009), and the broader bifactor-ESEM framework (Morin et al., 2016a, 2016b) has made it possible to benefit from the advantages traditionally associated with CFA/SEM analyses when relying on EFA measurement. ESEM is the broad analytic framework through which EFA has finally been connected to CFA/SEM (Marsh et al., 2014). ESEM makes it possible to consider, in the same model, CFA factors, EFA factors, and observed continuous or categorical indicators and variables. The goal of this chapter is to introduce ESEM and bifactor-ESEM to the readers. Three sets of online supplements are also provided: (a) Technical Supplements including annotated syntax files for all models estimated in this chapter; (b) Conceptual Supplements covering more advanced issues; (c) The two simulated data sets used in this chapter.

PSYCHOMETRIC MULTIDIMENSIONALITY

Before getting into the nuts and bolts of ESEM and bifactor-ESEM estimation, we first need to introduce these methods in a more conceptual manner to highlight the theoretical meaning of the analytic components included in these types of models. To this end, we need to introduce the concept of psychometric multidimensionality (Morin et al., 2016a, 2016b). Psychometric multidimensionality refers to the fact that factor indicators tend to truly reflect more than one "*thing*." According to classical test theory (CTT; e.g., Nunnally & Bernstein, 1994), any observed score (σ^2_{total}) is assumed to reflect two components: true score variance (σ^2_{true}) and random measurement error (σ^2_{error}), so that $\sigma^2_{total} = \sigma^2_{true} + \sigma^2_{error}$. These components lead to the definition of reliability (r_{xx}) as the ratio of true score variance on total variance: $r_{xx} = \sigma^2_{true}/\sigma^2_{total}$. Furthermore, σ^2_{true} itself is assumed to incorporate two components reflecting construct-irrelevant and construct-relevant sources of σ^2_{true}, the second of which is related to the concept of validity. This distinction is focused on the assessment of a single construct. When more than one construct is assessed, this perspective must be extended to consider a third source of σ^2_{true} related to the assessment of the other constructs. This phenomenon has been labelled construct-relevant psychometric multidimensionality by Morin et al. (2016a, 2016b), and contrasted with construct-irrelevant psychometric multidimensionality, which refers to sources of true score variance related to methodological artefacts (e.g., wording effects). The former needs to be explicitly accounted for, whereas the latter needs to be controlled for. Although the need to control for construct-irrelevant multidimensionality has long been recognized (e.g., Marsh et al., 2010), construct-relevant multidimensionality is usually ignored in CFA/SEM as part of the conditional independence assumption. According to this assumption, the latent factors should be sufficient to explain the covariance among all indicators, while allowing for the incorporation of methodological controls. Indicators are thus expected to be related to only one factor, leaving their possible associations with other factors to be considered as sources of measurement errors left to be absorbed in other parts of the model. This error propagation phenomenon explains why CFA/SEM tends to result in inflated estimates of factor correlations (Asparouhov et al., 2015).

CONSTRUCT-RELEVANT PSYCHOMETRIC MULTIDIMENSIONALITY

Conceptually-Related Constructs

Most indicators on which we rely to measure latent constructs are imperfect and almost never provide a perfect reflection of a single factor. With multidimensional measures including conceptually-related subscales, items can be expected to present construct-relevant associations with more than one factor. Recognizing that assessing conceptually-related constructs requires the estimation of cross-loadings does not imply that the definition of our constructs lacks clarity. For example, an item designed to measure the sport self-concept (e.g., "I am good at sports") is likely to present weaker, but meaningful, relations with other conceptually-related constructs (e.g., athleticism is known to play a role in popularity and physical appearance). Whereas these construct-relevant associations with non-target constructs are routinely ignored in CFA/SEM, they are explicitly modelled as cross-loadings in EFA/ESEM. Acknowledging that our indicators might be imperfect does not mean that we should stop trying to improve our measures, but highlights the need to account for this imperfect nature via cross-loadings even if they remain small in magnitude and in the absence of a theoretical rationale able to support all of them.

The assessment of conceptually-related constructs is a critical and necessary prerequisite to ESEM applications. Thus, all constructs forming a single set of ESEM factors (i.e., involving loadings and cross-loadings between all indicators and all factors) should be conceptually related to one another. Whereas cross-loadings are often required and justified between the dimensions of a single instrument, they should not necessarily be included across instruments, although distinct questionnaires may sometimes be used to represent multiple conceptually-related facets in a way that requires cross-loadings. For instance, Morin et al. (2016c, 2017) relied on ESEM to represent the structure of different facets of psychological health and well-being measured using distinct dimensions. In contrast, the dimensions of a single questionnaire may sometimes tap into qualitatively distinct psychological processes among which cross-loadings would be inappropriate (e.g., social support from work colleagues and family members). When cross-loadings need to be included, then they should only be included across constructs located at the same position in the predictive model under investigation, and among constructs measured at the same occasion. Incorporating cross-loading between variables located at different stages of a theoretical "causal" chain would create a paradoxical non-recursive situation where the same indicator would define two constructs specified as predicting one another. ESEM makes it possible to integrate different sets of EFA factors in the same model (Marsh et al., 2020), with cross-loadings allowed within, but not across, the different sets.

Hierarchically-Ordered Constructs

Construct-relevant psychometric multidimensionality also occurs with hierarchically-ordered constructs: When specific indicators are designed to reflect the specific facets (e.g., verbal, mathematical, and abstract reasoning) of a global construct (e.g., global intelligence). The optimal way to account for this form of psychometric multidimensionality is to rely on bifactor models (Reise, Chapter 18, this volume)[1]. Bifactor models allow scores on the indicators to directly define a global factor (G-factor), reflecting the variance shared among all indicators, together with a series of non-redundant specific factors (S-factors) representing the variance shared among all indicators forming a subscale beyond that explained by the G-factor. Bifactor models are orthogonal (i.e., the S-factors are not correlated). This orthogonality makes it possible to partition the total covariance into these two non-redundant components. Bifactor solutions should be systematically considered whenever one has a theoretical or empirical reason supporting the idea that a series of dimensions might reflect a global overarching construct, while also retaining some degree of specificity (Morin et al., 2016c, 2017, 2020).

Bifactor-ESEM

The bifactor-ESEM framework provides a way to account for both sources of construct-relevant psychometric multidimensionality in the same model (Morin et al., 2016a, 2016b, 2020). When constructs are expected (logically, empirically, or theoretically) to incorporate both types of multidimensionality, researchers should rely on a systematic comparison of CFA, bifactor-CFA, ESEM, and bifactor-ESEM solutions to select the optimal representation of their data (Morin et al., 2016a, 2017, 2020). These four models are illustrated in Figure 1.

THE ESTIMATION OF ESEM AND BIFACTOR-ESEM MODELS

Estimation and Model Fit Assessment

ESEM is currently only available in Mplus (since version 5.1; Muthén & Muthén, 2021), although partial implementations exist in *R* (R Core Team, 2020) (Geiser, Chapter 13, this volume). We thus focus on Mplus. In Mplus, ESEM and bifactor-ESEM solutions can be estimated using the Maximum Likelihood Robust (MLR) estimator for models relying on continuous indicators, or the weighted least square estimator using a diagonal weight matrix (WLSMV) when relying on ordinal indicators (Moustaki, Chen, & Zhang, Chapter 8, this volume; Bovaird, Chapter 15, this volume). Given the robustness of MLR to multivariate non-normality and to multilevel nesting (e.g., students within classrooms) when used in combination with

[1] Due to their limitations, higher-order models are not recommended (see Section 1 of the Conceptual Supplements).

the Mplus design-based correction of standard errors (TYPE = COMPLEX; Asparouhov, 2005), there is no reason to favor ML when MLR is so readily available. However, for indicators following an ordinal response process, asymmetric response thresholds, and/or involving four or less response categories, WLSMV estimation should be favored (Finney & DiStefano, 2013).

ESEM and bifactor-ESEM models should be assessed using commonly recommended goodness-of-fit indices and interpretation guidelines (Marsh et al., 2005; West, Wu, NcNeish, & Avord, Chapter 10, this volume), which also hold for WLSMV estimation (Yu, 2002). More precisely, scores ≥ .95 and .90 on the comparative fit index (CFI) and on the TLI, or ≤ .06 and .08 on RMSEA respectively indicate excellent and acceptable fit to the data. In model comparisons (Preacher & Yaremych, Chapter 11, this volume), decreases in CFI/TLI ≥ .01 and increases in RMSEA ≤ .015 between a model and a more parsimonious one support their equivalence, and thus that the more parsimonious model should be retained (Chen, 2007)[2].

Rotation Procedures

Rotation procedures seek to simplify the interpretability of the factors estimated via EFA/ESEM and are required to achieve identification (Osborne, 2015). Most rotations are mechanical (i.e., do not incorporate input from the researcher regarding the expected factor structure), and can be organized along a continuum depending on whether they seek to minimize factor complexity (i.e., smaller within-factor variability of the factor loadings, greater factor correlations, smaller cross-loadings) or to minimize variable complexity (i.e., greater within-factor variability of the factor loadings, smaller factor correlations, greater cross-loadings). However, all rotations have identical covariance implications and can be considered to be equivalent (a phenomenon referred to as "rotational indeterminacy"). The choice of a rotation has an influence on the estimated factor correlations and cross-loadings, but irrespective of the procedure, rotated EFA/ESEM results always tend to provide a more accurate representation of the factors (and of their correlations) relative to CFA. Finally, whereas correlated-factors models (EFA/ESEM) should rely on oblique rotations (i.e., allowing the factors to be correlated; Preacher & MacCallum, 2003), bifactor-ESEM models should rely on orthogonal rotations to ensure the proper interpretation of the factors (Morin et al., 2016a, 2020).

Two rotation procedures dominate applied ESEM and bifactor-ESEM research. Geomin rotation is a mechanical procedure associated with an epsilon value that users can change to reduce the size of the cross-loadings or the size of the factor correlations. Geomin rotation can accommodate bifactor estimation. Users are advised (Marsh et al., 2009; Morin et al., 2013) to rely on an epsilon value of .5 to maximally reduce factor correlations, and thus obtain more accurate (i.e., less multicollinear) estimates of relations between constructs. Target rotation is a non-mechanical procedure guided by one's theoretical *a priori*. Target rotation is thus the recommended procedure for confirmatory applications of ESEM and bifactor-ESEM (Morin et al., 2020). In its most typical form, target rotation relies on the *a priori* specification of the main indicators of each construct, allowing the loadings to be freely estimated, but "targeting" all cross-loadings to be close to zero as possible while allowing them to be freely estimated. Setting these "targets" does not force these loadings or cross-loadings to take that specific value, but simply relies on this information to guide the rotation. With target rotation, informed "targets" can be specified for loadings and cross-loadings. Although informed targets can be used to request a very precise solution, they can also be used to help converge on more interpretable, or useful, results[3]. Statistical research has supported the value of informed

[2] The chi-square (χ^2) and χ^2 difference tests should also be reported for purposes of transparency. However, these tests should never be interpreted due to their sample size dependency and oversensitivity to minor (substantively unimportant) misspecifications (Marsh et al., 2005). We also do not advocate the standardized root mean square residual (SRMR), due to its sample-size dependency and unstable performance (Chen, 2007; Marsh et al., 2005).

[3] For instance, the meaning of one factor can "flip" as a result of the initial rotation (i.e., with negative loadings). In this situation, setting a few positive "targets" for the main loadings corresponding to their estimated value can help to "flip back" the factor. Similarly, researchers may expect loadings on a G-factor to reflect a continuum (ranging from high and positive for indicators located at one end of the continuum to high and negative for indicators located at the other end of the continuum). Although their results might reflect a continuum, this continuum might be centered at a different point (e.g., the middle indicators might have the highest factor loadings). Setting some additional targets may help to rotate the factor structure in a way that better reflects the hypothetical continuum.

targets when they are consistent with the true population model, while highlighting the risk associated with erroneous targets (Guo et al., 2019; Myers et al., 2013, 2015). We recommend the basic approach to target rotation (i.e., targeting all cross-loadings to be as close to zero as possible). When necessary, researchers can also use a limited number of targets (as accurate as possible) to help achieve a solution that is maximally interpretable. Researchers seeking to rely on a more informed set of theoretically-driven targets should be able to clearly document the relevance of these targets, and demonstrate their impact on the solution relative to the more basic approach advocated here.

Basic Specification

Let us imagine a model including two factors (F1 and F2), each defined by four items (X1 to X4, and Y1 to Y4). The basic ESEM solution using an oblique mechanical rotation (Geomin with an epsilon value of .5) would be specified in the following manner in Mplus:

```
ANALYSIS:
ESTIMATOR = MLR; ROTATION = GEOMIN (.5);
MODEL:
F1-F2 BY X1 X2 X3 X4 Y1 Y2 Y3 Y4 (*1);
```

"GEOMIN (.5)" could be replaced by any other mechanical rotation procedure. "F1-F2" indicate that 2 factors will be estimated. These could alternatively be specified as "F1 F2," "Factor1 Factor2," or any other name. Then, the "(*1)" appearing at the end indicates that these factors are estimated from this list of indicators as a single set of ESEM factors (involving all possible loadings and cross-loadings between these indicators and the two factors). Should one want to incorporate a second set of ESEM factors (labelled F3 and F4), estimated from a second series of indicators (W1 to W4 and Z1 to Z4), then the syntax would be expanded as:

```
ANALYSIS:
ESTIMATOR = MLR; ROTATION = GEOMIN (.5);
MODEL:
F1-F2 BY X1 X2 X3 X4 Y1 Y2 Y3 Y4 (*1);
F3-F4 BY W1 W2 W3 W4 Z1 Z2 Z3 Z4 (*2);
```

In this second example, a first set of ESEM factors (F1 and F2) is estimated using indicators X1 to X4 and Y1 to Y4, with all possible loadings and cross-loadings freely estimated between this first set of indicators and factors F1 and F2. Then, a second set of ESEM factors (F3 and F4), identified by "(*2)", is estimated from a second set of indicators (W1 to W4 and Z1 to Z4), with all loadings and cross-loadings freely estimated between this second set of indicators and factors F3 and F4. No cross-loading is estimated between F3-F4 and items X1 to X4 and Y1 to Y4, just like no cross-loading is estimated between F1-F2 and items W1 to W4 and Z1 to Z4.

To specify the same two sets of ESEM factors using target rotation, a more complete definition of each factor is required, and all cross-loadings should be targeted (using the "~" symbol) to have a value as close to 0 as possible ("~0"), leading to:

```
ANALYSIS:
ESTIMATOR = MLR; ROTATION = Target;
MODEL:
F1 BY X1 X2 X3 X4 Y1~0 Y2~0 Y3~0 Y4~0 (*1);
F2 BY Y1 Y2 Y3 Y4 X1~0 X2~0 X3~0 X4~0 (*1);
F3 BY W1 W2 W3 W4 Z1~0 Z2~0 Z3~0 Z4~0 (*2);
F4 BY Z1 Z2 Z3 Z4 W1~0 W2~0 W3~0 W4~0 (*2);
```

The lack of targets refers to the main factor loadings, whereas the cross-loadings are ascribed a target value of 0. Two sets of factors are requested using "(*1)" and "(*2)".

To estimate a bifactor-ESEM solution, one need to select an orthogonal bifactor rotation and to request the estimation of a global factor (G-factor). So, assuming a model including three specific factors (SF1 to SF3) and one global factor (GF) defined from indicators X1 to X4 (SF1), Y1 to Y4 (SF2), and Z1 to Z4 (SF3) and forming a single set of factors:

```
ANALYSIS:
ESTIMATOR = MLR; ROTATION = BI-GEOMIN (orthogonal .5);
MODEL:
GF SF1-SF3 BY X1 X2 X3 X4 Y1 Y2 Y3 Y4 Z1 Z2 Z3 Z4 (*1);
```

Or, with target rotation:

```
ANALYSIS:
ESTIMATOR = MLR; ROTATION = target (orthogonal);
MODEL:
SF1 BY X1* X2 X3 X4 Y1~0 Y2~0 Y3~0 Y4~0 Z1~0 Z2~0 Z3~0 Z4~0 (*1);
SF2 BY Y1* Y2 Y3 Y4 X1~0 X2~0 X3~0 X4~0 Z1~0 Z2~0 Z3~0 Z4~0 (*1);
SF3 BY Z1* Z2 Z3 Z4 X1~0 X2~0 X3~0 X4~0 Y1~0 Y2~0 Y3~0 Y4~0 (*1);
GF BY X1 X2 X3 X4 Y1 Y2 Y3 Y4 Z1 Z2 Z3 Z4 (*1);
```

Measurement Model Comparisons

The first step in the application of ESEM or bifactor-ESEM involves the estimation of the alternative measurement models illustrated in Figure 1 to determine the optimal representation of the data (for a more generic coverage of model comparisons, see Preacher & Yaremych, Chapter 11, this volume). The need to contrast these four models comes from their ability to absorb unmodelled sources of construct-relevant multidimensionality while retaining a satisfactory (and potentially similar) level of fit (Morin et al., 2016a). Indeed, excluding cross-loadings can result in inflated factor correlations in CFA (Asparouhov et al., 2015) or in inflated G-factor loadings in bifactor-CFA (Morin et al., 2016a). Likewise, an unmodelled G-factor risks resulting in inflated CFA or ESEM factor correlations, or of ESEM cross-loadings (Morin et al., 2016a). These four models should thus be compared sequentially (Morin et al., 2016a, 2017, 2020).

First, the results from the CFA (Figure 1a) and ESEM (Figure 1b) solutions should be contrasted to verify the presence of multidimensionality due to the assessment of conceptually-related constructs. This comparison favors ESEM when: (a) the ESEM solution results in an improved (or at least equivalent) fit to the data; (b) the ESEM solution yields reduced factor correlations; (c) the ESEM solution includes small-to-moderate cross-loadings, or larger, easily explained cross-loadings (i.e., large and unexpected cross-loadings should lead to model revision); and (d) the ESEM solution results in well-defined factors. Observing multiple, moderate-to-large cross-loadings in ESEM may also suggest an unmodelled G-factor.

Second, the retained ESEM or CFA solution should be compared with the matching bifactor solution (Figure 1c: Bifactor-CFA; Figure 1d: Bifactor-ESEM). This comparison favors the bifactor solution when it results in: (a) an improved (or at least equivalent) fit to the data relative to the previously solution; (b) a well-defined G-factor (strong loadings, or loadings matching expectations); and (c) a subset of well-defined S-factors[4]. To illustrate this strategy, we rely on two simulated data sets (Leite, Bandalos, & Shen, Chapter 6, this volume).

Simulated Data Sets

Data 1 was simulated according to an ESEM specification (Figure 1b) with main loadings varying from .500 to .900, factor correlations of .200, .300, and .350, and cross-loadings ranging from 0 to .200 in absolute value (i.e., each factor had two cross-loadings of 0, two of .050, two of .100, and two of .200; half of these cross-loadings were negative). Indicators were simulated to be continuous with a mean of 0 and a SD of 1. We simulated a multigroup model (two groups of 5000), including one non-invariant intercept with a value of -.05 in Group 1 and .05 in Group 2.

Data 2 was simulated from a bifactor-ESEM specification (Figure 1d), and includes 10000 participants. The S-factors were specified with loadings varying from .400 to .700. The G-factor was

[4] Bifactor models do not need to result in S-factors that are all strongly defined (Section 2 of the Conceptual Supplements). Weaker S-factors control for residual specificities among a set of indicators once the variance explained by the G-factor is accounted for (Morin et al., 2016a, 2017). These S-factors indicate that these indicators mainly serve to define the global construct, and that scores on these indicators seldom deviate from scores on the global factor. Both conclusions have theoretical and practical implications (Morin et al., 2016a).

simulated loadings ranging from .400 to .800. Cross-loadings were simulated to be identical to those used in Data 1. Data 2 was simulated as a longitudinal model with two time points, with correlated uniquenesses between the matching indicators over time (Marsh, 2007) ranging from 0 to .150, and incorporating the same non-invariant intercept as in Data 1 (over time). The mean of the G-factor was specified as increasing from 0 (Time 1) to .8 (Time 2).

Measurement Model Comparisons: Data 1

The goodness-of-fit of the four alternative models is reported in Table 1, and parameter estimates are reported in the top section of Table 2[5]. Following the model comparison procedure outlined above, we first contrast the CFA and ESEM solutions. Whereas the ESEM solution has an excellent level of fit, the fit of the CFA is unacceptable. Both solutions result in well-defined factors (CFA λ = .516 to .885; M = .685; ESEM λ = .478 to .928; M = .681). The ESEM solution result in the estimation of cross-loadings that are mostly statistically significant (22/24), including seven cross-loadings between .100 and .199, and two cross-loadings higher than .200. Although the size of the correlations between factors 1 and 2 (CFA r = .259; ESEM r = .245) and between factors 1 and 3 (CFA r = .252; ESEM r = .203) is only slightly reduced in ESEM relative to CFA (consistent with the relatively small cross-loadings), the correlation between factors 2 and 3 is more markedly reduced in ESEM (r = .225) relative to CFA (r = .342). These results all support the ESEM solution, which we then retain for comparison with its bifactor counterpart. This bifactor-ESEM solution results in a level of fit and cross-loading pattern similar to the ESEM solution, but produces a weakly-defined G-factor (λ = .116 to .838; M = .363; only 4 factor loadings > .400), arguing in favor of retaining the ESEM solution for Data 1.

Measurement Model Comparisons: Data 2

The goodness-of-fit of the four alternative models estimated for Data 2 using Time 1 responses is reported in Table 1. Parameter estimates appear in the bottom of Table 2. Whereas the ESEM solution has an excellent fit, the fit of the CFA is unacceptable. Both solutions result in well-defined factors (CFA λ = .702 to .930; M = .821; ESEM λ = .615 to .952; M = .780). The ESEM solution reveals statistically significant cross-loadings (24/24), including five cross-loadings between .100 and .199, one cross-loading between .200 and .299, and three cross-loadings higher than .300. Interestingly, the ESEM cross-loadings are higher for Data 2 than for Data 1, even though both population models included identical cross-loadings. This observation illustrates that cross-loadings may suggest the need to incorporate a G-factor. Finally, the size of the correlation between factors 1 and 2 (CFA r = .535; ESEM r = .427), 1 and 3 (CFA r = .579; ESEM r = .480), and 2 and 3 (CFA r = .589; ESEM r = .425) is reduced in ESEM relative to CFA. As a result, the ESEM solution is retained for comparison with its bifactor counterpart. The bifactor-ESEM solution results in a higher level of fit than the ESEM solution (ΔCFI = +.010; ΔTLI = +.020; ΔRMSEA = -.049) and in slightly smaller cross-loadings (ESEM cross-loadings $|\lambda|$ = .017 to .389, M = .116; bifactor-ESEM cross-loadings $|\lambda|$ = .002 to .242; M = .085). The bifactor-ESEM solution also reveals a well-defined G-factor (λ = .353 to .859; M = .616) and generally well-defined S-factors (λ = .256 to .740; M = .541) and is thus retained for Data 2.

Construct-Irrelevant Sources of Psychometric Multidimensionality

We presented ESEM and bifactor models as ways to account for the construct-relevant psychometric multidimensionality. We also noted that psychometric multidimensionality is not always construct-relevant. In many situations, the covariance among subsets of indicators is only partly explained by the factors included in the model, and the unexplained covariance is not related to the research question. Construct-irrelevant psychometric multidimensionality is often related to the wording of the indicators. Thus, incorporating positively-worded (e.g., I like working for my organization) and negatively-worded (e.g., I hate my job) indicators to reflect the same set of factors creates a methodological artefact (i.e.,

[5] The initial estimation of the bifactor-ESEM solution resulted in a negative estimate of the uniqueness of item Z2, forcing us to re-estimate this solution while constraining this uniqueness to be higher than 0 (as shown in the online supplements). We note that our conclusions would remain unchanged had we retained the first (improper) solution. Convergence issues like this one frequently happen when working with ESEM models, and this procedure can be considered as helping the rotation procedure to converge on the best proper solution.

indicators sharing a similar wording share commonalities unrelated to the constructs; Marsh et al., 2010). Parallel wording (e.g., I like working with my supervisor; I like working with my colleagues) result in a similar artefact (Marsh et al., 2013a). Parallel wording is critical to consider in longitudinal studies where the same indicators are administered over time if one wants to avoid converging on inflated stability estimates (Marsh, 2007). When constructs are assessed using a mixture of informants (e.g., self, parent, and teacher reports) or methods (e.g., questionnaires, interviews), the ratings from each informant/method share commonalities that are not relevant to the constructs (Eid et al., 2008).

Construct-irrelevant sources of psychometric multidimensionality need to be controlled to ensure that it is not absorbed in other parts of the model. This control can be achieved by adding correlated uniquenesses among the relevant indicators, or by an orthogonal method factor reflecting the variance shared between these indicators. These approaches are illustrated in the top of Figure 2 (2a and 2b). The global factor was placed at the top to maximize clarity, and can be added, or removed and replaced by factor correlations depending on whether one wants to estimate an ESEM or bifactor-ESEM solution. Cross-loadings can also be removed to obtain a CFA or bifactor-CFA solution. In Figures 2a and 2b, items X1, Y1, and Z1 are negatively-worded. Correlated uniquenesses can be added to account for this wording effect (in bold):

```
MODEL:
SF1 BY X1* X2 X3 X4 Y1~0 Y2~0 Y3~0 Y4~0 Z1~0 Z2~0 Z3~0 Z4~0 (*1);
SF2 BY Y1* Y2 Y3 Y4 X1~0 X2~0 X3~0 X4~0 Z1~0 Z2~0 Z3~0 Z4~0 (*1);
SF3 BY Z1* Z2 Z3 Z4 X1~0 X2~0 X3~0 X4~0 Y1~0 Y2~0 Y3~0 Y4~0 (*1);
GF BY X1 X2 X3 X4 Y1 Y2 Y3 Y4 Z1 Z2 Z3 Z4 (*1);
X1 WITH Y1; X1 WITH Z1; Y1 WITH Z1;
```

Alternatively, an orthogonal method factor can be specified as:

```
MODEL:
SF1 BY X1* X2 X3 X4 Y1~0 Y2~0 Y3~0 Y4~0 Z1~0 Z2~0 Z3~0 Z4~0 (*1);
SF2 BY Y1* Y2 Y3 Y4 X1~0 X2~0 X3~0 X4~0 Z1~0 Z2~0 Z3~0 Z4~0 (*1);
SF3 BY Z1* Z2 Z3 Z4 X1~0 X2~0 X3~0 X4~0 Y1~0 Y2~0 Y3~0 Y4~0 (*1);
GF BY X1 X2 X3 X4 Y1 Y2 Y3 Y4 Z1 Z2 Z3 Z4 (*1);
MF BY X1* Y1 Z1; MF@1;
MF WITH GF@0 SF1@0 SF2@0 SF3@0;
```

Schweizer (2012) warned that *ex-post-facto* correlated uniquenesses should be avoided. We agree and reinforce that the approaches described in this section should be implemented in an *a priori* manner. Method factors have the advantage of resulting in a direct, explicit, and interpretable estimate of construct-irrelevant sources of variance. However, they also bring more complexity to the model, and more often result in convergence problems, especially in applications of ESEM and bifactor-ESEM due to the complexity of these models (i.e., cross-loadings and rotation). When multiple sources of construct-irrelevant multidimensionality need to be controlled (e.g., using a mixture of positive and negative items with parallel wording), it may be difficult to incorporate method factors for all sources. In this case, method factors are more naturally suited to negative-wording, and correlated uniquenesses to parallel-wording. When more than one method factor is included, it is important to leave method factors linked to different types of multidimensionality (e.g., parallel *vs* negative) uncorrelated with one another.

When indicators are rated by different informants or come from different methods, the situation becomes more complex, as each type of informant or method is tied to one source of construct-irrelevant psychometric multidimensionality (for details, see Eid, Koch and Geiser, Chapter 19, this volume). This situation typically calls for correlated traits correlated methods (CTCM) models, in which one method factor is added to account for each type of informant (or method). In these models, reports provided by distinct informants (or methods) may share some commonality not explained by the constructs, suggesting the need to incorporate correlations between the method factors. For example, parental reports may share something with teachers' reports that is not shared by self-reports (i.e., an adult perspective), self-reports may share something with parental reports that is not shared by teachers' reports (i.e., a household perspective), and self-reports may share something with teachers' reports that is not shared by parental reports (i.e., a school perspective). A CTCM model is illustrated in Figure 2c, where X1, X2, Y1, Y2, Z1, Z2 are self-reported, X3, Y3, and Z3 are reported by the teacher, and X4, Y4, and Z4 are reported by the parents. Unfortunately, this type of model tends to converge on improper solutions, or not to converge, especially in ESEM and bifactor-ESEM applications. This tendency is not surprising, as the indicator-level covariance serves to estimate many different sources of covariance (the trait factors, cross-loadings, bifactor, method factors, correlations between the method factors, and correlations between the trait factors).

To address this limitation, Eid et al. (2008) proposed to remove one method factor. The resulting correlated trait correlated method minus one model, or CTC(M-1), is illustrated in Figure 2d. In this example, the method factor associated with the self-reports has been removed. Removing this factor "anchors" the definition of the trait factors into the method/informant associated with the omitted method factor. The main factors reflect self-reports and what they share with other types of reports. The remaining method factors reflect the unique nature of these other reports (e.g., how parental and teachers' reports differ from self-reports). With CTCM or CTC(M-1) models, it is possible to incorporate correlated uniquenesses or other method factors to control for other forms of multidimensionality. In these cases, no correlations should be included between method factors linked to different controls. No cross-loadings should be estimated between method factors. CTCM and CTC(M-1) models can be specified as:

```
MODEL:
SF1 BY X1* X2 X3 X4 Y1~0 Y2~0 Y3~0 Y4~0 Z1~0 Z2~0 Z3~0 Z4~0 (*1);
SF2 BY Y1* Y2 Y3 Y4 X1~0 X2~0 X3~0 X4~0 Z1~0 Z2~0 Z3~0 Z4~0 (*1);
SF3 BY Z1* Z2 Z3 Z4 X1~0 X2~0 X3~0 X4~0 Y1~0 Y2~0 Y3~0 Y4~0 (*1);
GF BY X1 X2 X3 X4 Y1 Y2 Y3 Y4 Z1 Z2 Z3 Z4 (*1);
M1 BY X1* X2 Y1 Y2 Z1 Z2; M1@1;! Remove this line for CTC(M-1)
M2 BY X3* Y3 Z3; M2@1;
M3 BY X4* Y4 Z4; M3@1;
M1 WITH GF@0 SF1@0 SF2@0 SF3@0; ! Remove this line for CTC(M-1)
M2-M3 WITH GF@0 SF1@0 SF2@0 SF3@0;
```

Measurement Invariance

Tests of measurement invariance (Millsap, 2011; also see Widaman & Olivera-Aguilar, Chapter 20, this volume) are critical to the assessment of construct validity and generalizability of scores obtained on a specific measure across types of participants or measurement occasions. Measurement invariance is a prerequisite to unbiased comparisons of scores obtained across these types of participants or occasions. Tests of measurement invariance are conducted sequentially, starting from a model of *configural* invariance (i.e., the same measurement model, same number of factors, same indicators, and same specifications of indicators-to-factors associations). This model provides a baseline of comparison for subsequent models. A lack of configural invariance, as indicated by a lack of model fit, makes it irrelevant to pursue any other tests of measurement invariance and indicates that measurement properties are not comparable across groups or occasions. However, assuming that the model retained for tests of invariance achieves a satisfactory fit to the data in the total sample, as well as across groups or occasions, this model can still be flagged as ill-fitting by the RMSEA and the TLI due to their correction for parsimony, particularly in ESEM (due to the number of estimated loadings and cross-loadings). An adequate level of fit on the CFI for the configural model — when coupled with satisfactory fit on the TLI and RMSEA for the next model of weak invariance, and no substantial decrease in model fit between these two models for the CFI, TLI, and RMSEA — is thus sufficient to support the *weak* invariance of the solution (equality of the factor loadings across groups or occasions). Weak invariance is prerequisite to all further tests of invariance, and to unbiased comparisons of relations among constructs or of construct variability across groups or occasions.

As indicated by Widaman & Olivera-Aguilar (Chapter 20, this volume), the next steps involves tests of strong (equality of item intercepts for continuous indicators or of response thresholds for ordinal indicators, which is a prerequisite to tests of latent mean differences), strict (equality of the item uniquenesses, which is a prerequisite to comparisons involving observed scores but also a desiragble property as it increases parsimony), latent variance-covariance (equivalence of the factor variances and covariances), and latent means (equivalence of the factor means) invariance across groups or occasions. Tests of measurement invariance can be complemented by tests of predictive invariance (ideally starting from a model of strict, or at least strong, invariance), where the regression slopes, intercepts (i.e., the mean of the outcome) and residual (i.e., the variance of the outcome) are progressively constrained to equivalence.

Invariance is not an an all or none issue. Partial invariance of a majority of indicators of each factor remains sufficient for unbiased comparisons of latent variances, covariances, means, or regressions (Byrne et al., 1989). However, with ESEM or bifactor-ESEM models, it is currently not possible to test for the partial invariance of a subset of factor loadings/cross-loadings, variances, covariances, or means. Likewise, although it is possible to separately test for the invariance of the latent variances and covariances (although the former needs to be tested before the latter) in CFA/SEM, it is not possible to divide these tests in ESEM or bifactor-ESEM models. However, we later present a solution to these limitations.

The Multigroup Approach: Illustration[6]

The results from the tests of measurement invariance of the retained ESEM solution across the two simulated groups of participants (Data 1) are reported in Table 1. The results first indicate that the configural model has an excellent level of fit, supporting the adequacy of this model across groups. The fit indices remain essentially unchanged for the model of weak invariance, supporting the equivalence of the factor loadings across groups. However, the model of strong invariance results in a substantial decrease in model fit (ΔCFI = -.070; ΔTLI = -.090; ΔRMSEA = +.070), indicating a lack of intercept invariance. The model modification indices (obtained by requesting "OUTPUT: MODINDICES;" and reported in the output under "MODEL MODIFICATION INDICES") associated with this failed solution indicate that this lack of invariance is primarily due to item Z2, whose intercept (indicated by [Z2]) is associated with the largest modification index[7]. Equality constraints were thus removed for this intercept, leading to a model of partial strong invariance (Byrne et al., 1989), which is supported by the data (i.e., fit comparable to that of the weak invariance solution). Starting from this model, the next models of strict, latent variance-covariance, and latent means invariance are all supported by the data.

The Longitudinal Approach: Illustration

The results from the tests of measurement invariance of the retained bifactor-ESEM solution across the two simulated time points (Data 2) are reported in Table 1. The configural model results in an excellent fit, supporting its adequacy. The fit indices are unchanged for the model of weak invariance, supporting the equivalence of the factor loadings. The model of strong invariance results in a substantial decrease in model fit (ΔCFI = -.016; ΔTLI = -.022; ΔRMSEA = +.038), indicating a lack of intercept invariance. The modification indices of this failed solution and the parameter estimates from the previous model of weak invariance both indicate that the lack of invariance is primarily due to item Z2. Equality constraints were thus removed from this intercept, leading to a model of partial strong invariance which is supported by the data. Starting from this model, the next models of strict and latent variance-covariance invariance are also supported by the data, but not the model of latent mean invariance (ΔCFI = -.013; ΔTLI = -.015; ΔRMSEA = -.030). As it is currently impossible to test the partial invariance of latent means in ESEM and bifactor-ESEM, we retain the model of latent variance-covariance invariance as our final model. Examination of the parameter estimates of this model reveals that, with the Time 1 latent means fixed to 0 for identification

---

[6] The logic of invariance testing is the same for CFA and ESEM (see Widaman & Olivera-Aguilar, Chapter 20, this volume). However, the syntax is specific to ESEM, and the limitations of ESEM makes these tests slightly more complex than in CFA. Therefore, we provide a complete illustration, with accompanying MLR and WLSMV syntax in the Technical Supplements. The results reported in the main text are based on MLR.

[7] Using modification indices to locate a partially invariant solution should be done in a stepwise manner. First the item with the largest modification index should be located. Invariance constraints on this item should then be relaxed. If the fit of the resulting model matches that of the last retained model, then the process can stop. Otherwise, the process should be repeated, freeing one parameter at a time. The examination of the modification indices of the failed solution should be supplemented by an examination of the freely estimated parameters from the last supported solution. Here, the free intercepts from the weak invariance solution also flagged item Z2.

purposes and all latent variances fixed to 1, the Time 2 latent means (freely estimated in SD units as deviations from Time 1 means) are slightly higher than the Time 1 means for the S-factor 1 (+.079 SD, $p \leq .01$), the S-factor 3 (+.061 SD, $p \leq .01$) and the G-factor (+.764 SD, $p \leq .01$), but lower than the Time 1 mean for the S-factor 2 (-.125 SD, $p \leq .01$). Arguably, observing small, and yet statistically significant, differences linked to the S-factors (which we know not to be part of our population model) could simply reflect the large sample size of this simulated data set (N = 10,000).

Tests of Differential Item Functioning: Illustration

There are situations where the full taxonomy of tests of measurement invariance cannot be realistically applied: (a) When group-specific sample sizes are too small; (b) when multiple grouping variables (and their interactions) have to be considered; or (c) when testing for measurement biases occurring as a function of continuous variables (e.g., age, salary, pretest scores) which should not be recoded into a smaller number of discrete groups (to avoid the reduction in precision and power that accompanies the categorization of continuous variables). Due to the greater complexity of ESEM and bifactor-ESEM models, these types of issues tend to be more frequent than with CFA. In these situations, tests of differential item functioning (DIF) can be conducted using multiple indicators multiple causes (MIMIC) models (e.g., Muthén, 1989). These models involve the addition of an observed predictor to the previously retained measurement model. Tests of DIF (more precisely of monotonic DIF) correspond to tests of the invariance of the intercepts (or response thresholds with WLSMV) through the verification of whether the effect of the predictor on the item responses can be captured entirely by its effect on the factors, or whether it also influences item response beyond its impact on the factors. These tests involve three alternative models (e.g., Morin et al., 2013), which are illustrated in Figure 3. In this Figure, the cross-loadings and G-factors have only been removed to maximise clarity.

The *null effects* model assumes that the predictor(s) have no effect on the factors (including the G-factor in bifactor models) and item responses (i.e., all dashed paths from Figure 3 are incorporated but constrained to be 0). The *saturated* model involves the free estimation of all paths linking the predictor(s) to item responses (i.e., the dashed greyscale paths), while keeping the effects of the predictor(s) on the factors constrained to 0. The *invariant* model involves the free estimation of all paths linking the predictor(s) to the factors (i.e., the dashed black paths), while keeping the effects of the predictor(s) on the item responses constrained to 0. Comparing the null effects and saturated models tests whether the predictors influence item responses. When this is the case, comparing the saturated and invariant models tests whether this influence can (if both models have a similar fit) or not (if the saturated model fits better) be fully explained in terms of their association with the factors. When the saturated model fits better than the invariant model, then there is evidence of monotonic DIF (i.e., non-invariance of the item intercepts or response thresholds). In this case, it might be appropriate to investigate models of partial DIF.

We illustrate this approach using Data 1 using the grouping variable as the predictor. The fit of the three alternative models is reported in the bottom of Table 1. Although the null effects model results in a satisfactory fit according to the CFI, the RMSEA value is barely acceptable, and the TLI flags this model as ill-fitting. The saturated model has an excellent fit, indicating that there are at least some effects of the grouping variable on the item responses. However, the invariant model results in a level of fit comparable to that of the null effects model, and much lower than that of the saturated model, suggesting that the effects of the grouping variable do not occur at the level of the factors, thus providing early evidence of DIF. The parameter estimates associated with the saturated model and the modification indices associated with the invariant model suggests that this DIF is limited to item Z2. As a result, we estimated a model of partial invariance, allowing the groups to predict scores on the latent factors and responses on item Z2. This model is associated with fit equivalent to that of the invariant model and is thus retained. Examination of the parameter estimates from this model reveals that the groups share no relation with the latent factors (Factor 1: b = -.003, $p > .05$; Factor 2: b = -.016, $p > .05$; Factor 3: b = .020, $p > .05$), but that members of the second group tended to score, on the average, 1 SD higher on item Z2 than members of group 1 with similar scores on the latent factors (b = .999, $p \leq .01$).[8]

[8] Marsh et al. (2013b) proposed a multi-group hybrid MIMIC model to test the extent to which the categorisation of a

Predictive Models

When comparing ESEM and bifactor-ESEM to traditional CFA/SEM models, the main difference concerns the incorporation of an EFA measurement structure to the solution. This is why the present chapter mainly focuses on the measurement component of these models. Once the optimal measurement structure has been identified, one can easily add additional observed variables and latent factors corresponding to all four models illustrated in Figure 1 (and including any or all of the methodological controls illustrated in Figure 2), before converting the measurement model into a any of the predictive models discussed in the other chapters from this volume. In Mplus, the specification of predictive paths is done using the ON function, such as:

```
MODEL:
F1 BY X1* X2 X3 X4 Y1~0 Y2~0 Y3~0 Y4~0 (*1);
F2 BY Y1* Y2 Y3 Y4 X1~0 X2~0 X3~0 X4~0 (*1);
F3 BY W1 W2 W3 W4 Z1~0 Z2~0 Z3~0 Z4~0 (*2);
F4 BY Z1* Z2 Z3 Z4 W1~0 W2~0 W3~0 W4~0 (*2);
F3-F4 ON F1-F2;
```

In prediction, one must keep in mind a key limitation of ESEM and bifactor-ESEM: All factors forming a single set should be related in the same manner to other variables. In the above example where F3 and F4 are regressed on F1 and F2, it would not be possible to estimate a model in which only F3 is regressed on F1, and in which only F4 is repressed on F2.

LIMITATIONS AND SOLUTIONS

The need to rely on a rotation procedure is the reason why it took so much time to connect EFA to the CFA/SEM framework, and the reason why this connection is still incomplete. More precisely, some limitations remain relative to CFA applications:

(a) It is impossible to test the partial invariance of the factor loadings/cross-loadings, latent variances/covariances, and latent means;

(b) All factors forming a single set of ESEM or bifactor-ESEM factors need to be related in the same manner to other variables included in the model.

(c) It is impossible to estimate a second-order factor from a set of ESEM factors;

(d) It is impossible to estimate a latent curve or latent change models from longitudinal sets of ESEM or bifactor-ESEM factors due to the impossibility to constrain ESEM or bifactor-ESEM latent means;

(e) It is impossible to estimate bifactor-ESEM models including more than one G-factor while allowing these G-factors to be correlated with one another;

(f) In models including more than one set of ESEM factors or bifactor-ESEM factors, these various sets need to be estimated using the same rotation procedure, which makes it impossible to include in the same model one set of ESEM factors (oblique) with one set of bifactor-ESEM factors (orthogonal);

(g) It is impossible to estimate mixture or factor mixture models where the profiles are defined based on ESEM or bifactor-ESEM factors.

(h) One last limitation is specific to bifactor (CFA or ESEM) models. As noted by Koch et al. (2018), when orthogonal factors (i.e., with an unconditional covariance of 0) from a bifactor solution are used as endogenous variables (i.e., outcomes, regressed on other variables), only their residual covariances (rather than their unconditional ones) can be constrained to be zero. For this reason, the initial properties of the bifactor solution are lost. Unfortunately, the solutions proposed to Koch et al. (2018) to address this issue remain very complex, lack the flexibility normally associated with bifactor models (e.g., dividing the predictors and outcomes into global and specific components and estimating

---

continuous predictor results in a loss of information. While the MIMIC approach tests for monotonic DIF, it assumes—without testing it—the invariance of the loadings (non-monotonic DIF). Although approaches exist to test for non-monotonic DIF (Barendse et al., 2010), these approaches cannot be implemented with ESEM or bifactor-ESEM models, at least without relying on the ESEM-within-CFA approach presented later. Even then, this approach remains too computer-intensive for most applications relying on moderately complex measurement models.

predictions limited to a single type of component, or relying on a two-step residual procedure). The factor score approach described below provides a much simpler solution[9].

Before addressing more generalizable solutions, a simple solution can be used for (e): Using an orthogonal target rotation, one can estimate a set of ESEM factors (i.e., uncorrelated S-factors), together with correlated global factors defined using CFA representation:

```
ANALYSIS:
ESTIMATOR = MLR; ROTATION = target (orthogonal);
MODEL:
SF1 BY X1* X2 X3 X4 Y1~0 Y2~0 Y3~0 Y4~0 Z1~0 Z2~0 Z3~0 Z4~0 (*1);
SF2 BY Y1* Y2 Y3 Y4 X1~0 X2~0 X3~0 X4~0 Z1~0 Z2~0 Z3~0 Z4~0 (*1);
SF3 BY Z1* Z2 Z3 Z4 X1~0 X2~0 X3~0 X4~0 Y1~0 Y2~0 Y3~0 Y4~0 (*1);
GF1 BY X1* X2 X3 X4 Y1 Y2;
GF2 BY Y3* Y4 Z1 Z2 Z3 Z4;
GF1@1; GF2@1; GF1 WITH GF2*;
GF1 WITH SF1@0 SF2@0 SF3@0;
GF2 WITH SF1@0 SF2@0 SF3@0;
```

ESEM-within-CFA

Marsh et al. (2013b) proposed ESEM-within-CFA (EWC) as a simple and efficient solution to many of those limitations. Morin et al. (2013) provided multiple examples of this method (i.e., partial invariance of factor loadings/cross-loadings, factor variances, and factor covariances; test of latent mean differences relying on contrast codes; tests of mediation and indirect effects involving a subset of factors; latent change models). Essentially, EWC involves imposing the same number of restrictions (m^2 restrictions, where m = number of factors) imposed as part of the rotation procedure. In EWC, these restrictions are imposed by setting all factor variances to 1, and by selecting one referent indicator per factor and fixing all cross-loadings for this referent indicator to their value in the original ESEM solution (values of these parameters can be obtained in the "CFA MODEL COMMAND WITH FINAL ROTATED ESTIMATES USED AS STARTING VALUES" section of the output by using the "OUTPUT: SVALUES;" function). EWC models need to be built from an already existing ESEM solution, using the estimates from this initial solution as start values (specified with *). Although it is recommended to select referent indicators with a strong main loading and weak cross-loadings, any referent indicator can be used as long as the model is specified using the start values from the unconstrained model. The resulting EWC solution will then have the same degrees of freedom and, within rounding error, the same chi-square, goodness of fit statistics, and parameter estimates as the original solution, and can be used as the starting point for the remaining analyses. However, standard errors may be slightly inflated in the EWC solution (i.e., marginally significant results should be cautiously interpreted).

To construct an EWC solution corresponding to a bifactor-ESEM solution, the G-factor must be considered as any other factor. Thus, a different referent indicator needs to be selected for the G-factor, and the loadings of the other referent indicators on this G-factor should also be fixed. The factor correlations also need to be fixed to 0. This last constraint will produce an EWC solution that will differ from the original by a number of *df* corresponding to the number of correlations fixed to 0 (as these are rotated to 0 in bifactor-ESEM, rather than fixed).

EWC can be used to analyse higher-order ESEM solutions where the second-order factors are defined according to a CFA specification (to rely on an ESEM specification of the second-order factors, the next two approaches are required), with one difference. Whereas the traditional EWC approach fixes the factor

[9] Koch et al. (2018) also noted another problem, which can be solved in the same manner. However, this second problem is tied to a misunderstanding of bifactor models in which the authors assumed that the S-factors should necessarily have a mean of 0 (which becomes an intercept in predictive analyses) to maintain their interpretation as residual scores, an interpretation that is only relevant to CTCM models where the factors reflect residual variance linked to methodological artefacts, not to bifactor models where the S-factors are substantively meaningful in their own right and for which it is possible, and adequate, the estimate factor means.

variances to 1, in a higher-order solution these variances become residual variances (i.e., the disturbance of the first-order factor unexplained by the second-order factor). Therefore, in the EWC solution used to analyse higher-order factor structure, the main loading of the referent indicator should also be fixed to its ESEM value. Factor variances should also be given a start value of 1 rather than being fixed to 1 (Morin & Asparouhov, 2018).

In the online supplements, we provide an EWC input to replicate Data 1 ESEM solution. This solution results in a level of fit that is identical to that of the original solution. To illustrate the estimation of higher-order models, we provide an EWC equivalent to the non-retained ESEM solution estimated in Data 2 at Time 1. This solution also results in a level of fit that is identical to that of the original solution. Likewise, because higher-order factor solutions including one second-order factor and three first-order factors are mathematically equivalent to first-order solutions including three factors (i.e., involving the replacement of three first-order factor correlations by three second-order factor loadings), the fit of the EWC higher-order solution is also equivalent to that of the original solution. However, this solution reveals that the second-order factor is only moderately well-defined by its three indicators (λ = .615 to .691), which is consistent with the fact that the population model was simulated from a bifactor specification, but also with the superiority of the bifactor-ESEM solution in this specific context.

Finally, using Data 2, we provide an EWC replication of the longitudinal bifactor-ESEM model of invariance of the latent variance-covariance. The initial model results in a level of fit that almost identical to that of the original solution with a difference of 6 df related to the need to fix 6 time-specific covariances (invariant over time) to 0. From this model, we investigated whether the lack of latent mean invariance could be attributed entirely to the G-factor by estimating a model of partial latent mean invariance in which all S-factor means were set to be equal over time. This model of partial latent mean invariance resulted in a fit equivalent to that of the model of latent variance-covariance invariance, and was thus supported by the data. In this model, the mean of the G-factor was found to increase from a value of 0 at Time 1 to a value of .758 at Time 2 ($p \leq .01$). As a last example provided in the online supplements, we then convert this solution to a latent change model designed to specifically model change occurring over time in the G-factor. This solution is a simple re-expression of the previous one, resulting in a virtually identical level of fit to the data, and resulting in the estimation of a latent change factor (which can then be used in prediction) with a mean of .758 and a variance of 1.984.

Despite its advantages, EWC requires an initial ESEM or bifactor ESEM model, and thus will always have the same level of complexity as the original solution. In addition, EWC should only be used as a single intermediary step toward the final solution, as the further away one gets from the original solution, the least valid the EWC approximation will be. Thus, one should not use EWC to investigate the partial invariance of factor loadings, and then remain within EWC to test strong, strict, latent variance-covariance, and latent mean invariance.

Factor Covariance Matrix and Factor Scores

The complexity of ESEM and bifactor-ESEM might remain a problem with EWC (e.g., research including multiple measures or time points, tests of latent interactions, or person-centered analyses). In these situations, some might want to revert back to scale scores (i.e., the average or sum of the indicators forming each subscale). This solution might be justified for simple CFA models including highly reliable factors. However, scale scores present two key limitations: (a) they are uncorrected for unreliability; and (b) they fail to preserve the nature of the measurement model (i.e., the size of the factor loadings, the cross-loadings, the global/specific nature, and the controls for construct-irrelevant forms of multidimensionality).

A first alternative solution is to save the factor covariance matrix associated with the retained ESEM or bifactor-ESEM solution, and to use this covariance matrix in further analyses. This approach completely preserves the nature of the measurement model, including the correction for unreliability, but is limited by the need to incorporate all variables required for the main analyses in the model used to save this covariance matrix. This approach is also not relevant to person-centered analyses (Morin & Litalien, 2019). We have found this approach to be especially useful for higher-order models, particularly when the second-order structure also needs to be defined by an ESEM representation (multiple second-order factors and cross-loadings). The following command is added at the end of the model from which the factor covariance matrix is exported: "SAVEDATA: TECH IS FCOV.DAT;". A new data file, named "FCOV.DAT" (the name can be changed) is created and can used as the data in the next analysis: "DATA: FILE IS FCOV.DAT; TYPE IS MEANS COVARIANCE; NOBSERVATIONS = 500;" (where "500" is changed to reflect the sample size). In this new data set, variables are in the same order as in the initial model (details are provided in the TECH4 section of the output).

A more flexible solution is to save factor scores from the initial ESEM, bifactor-ESEM, or even CFA or bifactor-CFA solution, and to use these factor scores in further analyses (see Devlieger & Rosseel, Chapter 17, this volume). Factor scores preserve the nature of the measurement model (Morin et al., 2016c, 2017), afford researchers a partial correction for measurement error (Skrondal & Laake, 2001), and are particularly useful for person-centered analyses. Whenever the indicators used in person-centered analyses present a global/specific multidimensional structure (i.e., bifactor), ignoring this dual structure tends to result in the estimation of profiles presenting a low differentiation across indicators (Morin et al., 2016c, 2017). Using bifactor factor scores is the optimal way to account for this dual structure. However, one must keep in mind that these approaches, despite their flexibility, still only reflect a "patch," rather than a true alternative to fully latent models. To obtain factor scores, one adds the following at the end of the model from which the factor scores are exported: "SAVEDATA: FILE IS FSCORES.DAT; SAVE = FSCORES;", where "FSCORE.DAT" is the name of the data file including the factor scores (this name can be changed). The list, and order, of variables included in this data set appears at the end of the output in the "SAVEDATA INFORMATION" section. Variables excluded from the initial model can be included in this data set by listing them in the "AUXILIARY =" section of "DATA:".

COMMONLY VEHICULATED INACCURACIES

The fact that ESEM helps researchers achieve a more accurate representation of the latent factors, factor correlations, and factor regressions using all of the information available at the indicator level is now relatively well-established (Asparouhov et al., 2015; Mai et al., 2018; Morin et al., 2020). However, after living in the shadow of CFA/SEM for so long, residual misconceptions remain associated with ESEM (Morin et al., 2020). The most common misconceptions are that ESEM models lack of parsimony, lack of "simple structure," or lack of relevance for "confirmatory" applications. These misconceptions are particularly insidious as they seem to be anchored in logical propositions, and often emerge during the peer review process to create an unnecessary hurdle for less experienced researchers who thought (with reason) that they were using an improved form of statistical analysis.

Parsimony

By freely estimating all cross-loadings, ESEM is automatically less parsimonious than CFA models from which these cross-loadings are excluded. For this reason, when both types of models yield result in an equivalent representation of the data (i.e., when both result in a similar level of fit to the data and in similar factor correlations), CFA should theoretically be favored over ESEM. There is, thus far, nothing flawed in these affirmations. However, it has also been suggested that freely estimating all cross-loadings automatically results in inflated model fit estimates. This affirmation is flawed for two reasons. First, it ignores the fact that, among the many fit indices available to help researchers select an optimal solution, some include a correction for parsimony (Marsh et al., 2005), and can thus support for the superiority (or equivalence) of a CFA representation of the data relative to an ESEM representation when the cross-loadings are unnecessary. For this reason, ESEM advocates have noted the importance of the Tucker-Lewis Index (TLI) and the Root Mean Square Error of Approximation (RMSEA) when considering these types of models (e.g., Morin et al., 2013, 2020). In addition, the model comparison strategies typically advocated for contrasting alternative ESEM and CFA solutions highlight the critical role of the factor correlations, which directly indicate whether the cross-loadings have an impact on improving the factor definition (Morin et al., 2020).

Simple Structure, Cross-Loadings, and Factor Definition

One major misconception is that cross-loadings change the definition of the factors. This misconception is often tied to a misunderstanding of the concept of "simple structure" advocated by Thurstone (1947). As an early developer of EFA, Thurstone proposed the concept of simple structure to guide the development of rotation procedures that would maximise the accuracy and interpretability of the factors. This principle suggests that factors should be anchored in distinctive sets of indicators contributing primarily to the definition of one factor relative to the others, but was never formulated to suggest that cross-loadings should be omitted from a model. Returning to the idea that cross-loadings may contaminate the meaning of the factors, this proposition is doubly inaccurate. First, factor analyses involve reflective measurement models, which position the factors as the cause of the scores obtained on the indicators, rather than the opposite (e.g., It is because I am intelligent that I can solve complex mathematical problems). Cross-loadings thus only reflect the influence of the factor on the indicators, and are therefore unable to change the meaning of the factors. Rather, they simply allow the factors to be estimated using all of the information present in the indicators. Second, CTT makes it clear that the meaning of a construct lies in the way it associates with other constructs rather than in the way it associates with its indicators. The former represents the essence of analyses of validity. The latter is more intimately related to analyses of reliability. As noted above, the observation that measurement models incorporating cross-loadings result in a more accurate representation of the factors, their correlations, and the regressions among them (Asparouhov et al., 2015; Mai et al., 2018) is consistent with the idea that cross-loadings help improve the definition of a factor rather than contaminate it. It remains true that observing cross-loadings that are both large and hard to explain theoretically, logically, or empirically is indicative that something is not going as planned, and that researchers should re-examine their assumptions, their model, and their data. However, this does not indicate that the items are *tainting* the meaning of the factor.

Exploratory or Confirmatory Methods or Purposes

Referring to CFA/SEM and EFA/ESEM as exploratory or confirmatory has resulted in a semantically driven misconception. The difference between these two approaches is simply linked to their mathematical underpinnings, which either involves the estimation of factors defined using an *a priori* subset of indicators (CFA/SEM) or defined using all indicators but allowing them to empirically define one factor more than the others (EFA/ESEM). From this methodological difference, it is easy to take the extra step and assume that CFA/SEM should be limited to *confirmatory* studies, whereas EFA/ESEM should be reserved for *exploratory* studies. However, it is possible to use both approaches for confirmatory *or* exploratory purposes. On the one hand, EFA/ESEM can be used to verify whether the observed factor structure matches our expectations when no additional constraints are added (i.e., a more stringent test than when one only estimates *a priori* associations). On the other hand, the *a priori* CFA/SEM solution may fail to fit the data, leading researchers to rely on *ex post facto* (or exploratory) modifications to obtain a well-fitting solution. In this second situation, EFA/ESEM carries significant advantages: Whereas the modification indices used to “correct” an ill-fitting CFA solution are estimated in a stepwise manner (one parameter at a time), EFA/ESEM simultaneously estimate all possible cross-loadings in a single step, often revealing issues that would have been impossible to identify in CFA/SEM (Morin & Maïano, 2011). EFA/ESEM may thus be more naturally suited to exploration than CFA/SEM. However, this does not mean that CFA is more naturally suited to confirmatory investigations for two reasons. First, EFA/ESEM has been found to result in a more accurate factor representation (Asparouhov et al., 2015; Mai et al., 2018), which highlights the unrealism of assuming that all indicators can provide a perfect reflection of one, and only one, factor. Second, target rotation (Browne, 2001; Reise, 2012) makes it possible to rely on an *a priori* specification of ESEM factors. It is thus not surprising that most of applications of ESEM have pursued confirmatory, rather than exploratory, goals (Marsh et al., 2014).

Further Considerations and Misconceptions

As ESEM and bifactor-models are relatively new in the arsenal of applied researchers, it should not come as a surprise to observed that best practice recommendations related to these types of models are ever evolving, and still the object of various misconceptions and debates. In the Conceptual supplements, we address in more details three of these areas of uncertainty, debates, and misconceptions: (1) The orthogonality of bifactor models, vanishing S-factors, and the bifactor S-1 fiction (Section 2 of the Conceptual Supplements); (2) Reliability estimation (Section 3); and (3) Power analyses (Section 4).

CONCLUSIONS

ESEM is a connection between traditional EFA measurement models, and the broader CFA/SEM framework. As a result, ESEM makes it possible to rely on EFA measurement models as part of most research applications relying on latent variable models while capitalizing on most of the advantages afforded by the CFA/SEM framework that are the object of the other chapters of this volume. This simple summary hides a far more widespread contribution of ESEM to applied research. By creating this missing connection, ESEM has launched a whole new era of statistical research, which has made it possible to realize that by relegating EFA to the role of second-class citizen, we were very close to throwing out the baby with the bathwater. ESEM and the broader bifactor-ESEM framework make it possible to achieve a more realistic, and accurate, depiction of the construct-relevant psychometric multidimensionality present in most of our measures. By ignoring this multidimensionality, the traditional CFA approach is unable to achieve a proper definition of the latent constructs of interest and results in inflated estimates of factor correlations to correct for the lack of consideration of the conceptually-related and hierarchically-ordered nature of our constructs (Asparouhov et al., 2015; Morin et al., 2020). When converted to predictive SEM models, these inflated correlations result in biased predictions and multicollinearity (Mai et al., 2018). However, and pending further developments, the “link” provided by ESEM remains imperfect, although some solutions to these imperfect connections have been proposed in the present chapter. Likewise, as is the case for most new developments, areas of debates, misconceptions, and uncertainties remains. We believe that, as ESEM applications become more widespread, more integrated solutions to these limitations will be developed, and a broader consensus will emerge regarding best practices. In the meantime, we hope that this chapter will motivate applied researchers to adopt this new analytical framework to achieve a higher level of precision in their research, and reviewers to refrain from imposing debatable or dubious approaches and practices to authors.

REFERENCES

Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling*, *12*, 411-434.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397-438.

Asparouhov, T., Muthén, B., & Morin, A.J.S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances. *Journal of Management*, *41*, 1561-1577.

Barendse, M.T., Oort, F.J., & Garst, G.J.A. (2010). Using unrestricted factor analysis with latent moderated structures to detect uniform and non-uniform measurement bias: A simulation study. *AStA Advances in Statistical Analysis*, *94*, 117-127.

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

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466.

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement. *Structural Equation Modeling*, *14*, 464-504.

Eid, M., Nussbeck, F.W., Geiser, C., Cole, D.A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, *13*, 230-253.

Finney, S.J., & DiStefano, C. (2013). Non-normal and categorical data in structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds), *Structural Equation Modeling: A Second Course, 2nd edition* (pp. 439-492). Greenwich, CO: IAP.

Guo, J., Marsh, H.W., Parker, P.D., Dicke, T., Lüdtke, O., & Diallo, T.M.O. (2019). A systematic evaluation and comparison between exploratory structural equation modeling and Bayesian structural equation modeling. *Structural Equation Modeling, 26*, 529-556.

Koch, T., Holtmann, J., Bohn, J., & Eid, M. (2018). Explaining general and specific factors in longitudinal, multimethod, and bifactor models. *Psychological Methods*, *23*, 505-523.

Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing exploratory structural equation modeling and existing approaches for multiple regression with latent variables. *Structural Equation Modeling*, *25*, 737-749

Marsh, H.W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. Dans G. Tenenbaum & R. C. Eklund (Éds.), *Handbook of on Sport*

Psychology (3rd ed.). New York, NY: Wiley.

Marsh, H.W., Abduljabbar, A.S., Abu-Hilal, M.M., Morin, A.J.S., Abdelfattah, F., Leung, K.C., Xu, M.K., Nagengast, B., & Parker, P. (2013a). Factorial, convergent, and discriminant validity of timss math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology, 105*, 108-128.

Marsh, H.W., Hau, K., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J.J. McArdle (Eds.), *Contemporary psychometrics* (pp. 275-340). Mahwah, NJ: Erlbaum.

Marsh, H.W., Guo, J., Dicke, T., Parker, P.D., & Craven, R.G. (2020). Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and set-ESEM: Optimal balance between goodness of fit and parsimony. *Multivariate Behavioral Research, 55*, 102-119.

Marsh, H.W., Morin, A.J.S., 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.

Marsh, H.W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A.J.S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16*, 439-476.

Marsh, H.W., Nagengast, B., & Morin, A.J.S. (2013b). Measurement invariance of Big-Five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and La Dolce Vita effects. *Developmental Psychology, 49*, 1194-1218.

Marsh, H.W., Scalas, L.F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg self-esteem scale. *Psychological Assessment, 22*, 366-381.

McCrae, R.R., Zonderman, A.B., Costa, 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 & Social Psychology, 70*, 552-566.

Millsap, R.E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

Morin, A.J.S., Arens, A.K., & Marsh, H.W. (2016a). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling, 23*, 116-139.

Morin, A.J.S., Arens, A.K., Tran, A., & Caci, H. (2016b). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the composite scale of morningness: Construct-relevant multidimensionality. *International Journal of Methods in Psychiatric Research, 25*, 277-288.

Morin, A.J.S., & Asparouhov, T. (2018). Estimation of a hierarchical Exploratory Structural Equation Model (ESEM) using ESEM-within-CFA. Montreal, QC: Substantive Methodological Synergy Research Laboratory.

Morin, A J.S., Boudrias, J.-S., Marsh, H.W., Madore, I., & Desrumaux, P. (2016c). Further reflections on disentangling shape and level effects in person-centered analyses: An illustration aimed at exploring the dimensionality of psychological health. *Structural Equation Modeling, 23*, 438-454.

Morin, A.J.S., Boudrias, J.-S., Marsh, H.W., McInerney, D.M., Dagenais-Desmarais, V., Madore, I., & Litalien, D. (2017). Complementary variable- and person-centered approaches to the dimensionality of psychometric constructs: Application to psychological wellbeing at work. *Journal of Business and Psychology, 32*, 395-419.

Morin, A.J.S., & Litalien, D. (2019). Mixture modeling for lifespan developmental research. In *Oxford Research Encyclopedia of Psychology*. Oxford University Press.

Morin A.J.S., & Maïano C. (2011). Cross-validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM). *Psychology of Sport and Exercise, 12*, 540-554.

Morin, A.J.S., Marsh, H.W., & Nagengast, B. (2013). Exploratory structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 395-436). Greewich, Connecticut: IAP.

Morin, A.J.S., Myers, N.D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling and bifactor-ESEM. In G. Tenenbaum, & Eklund, R.C. (Eds.), *Handbook of Sport Psychology, 4th Edition, Vol. 2* (pp. 1044-1073). Wiley.

Muthén, B. (1989). Latent variable modeling in heterogenous populations. *Psychometrika*, *54*, 557-585.

Muthén, L.K., & Muthén, B. (2021). *Mplus user's guide*. Los Angeles: Muthén & Muthén.

Myers, N.D., Ahn, S., & Jin, Y. (2013). Rotation to a partially specified target matrix in exploratory factor analysis: How many targets? *Structural Equation Modeling*, *20*, 131-147.

Myers, N.D., Jin, Y., Ahn, S., Celimli, S., & Zopluoglu, C. (2015). Rotation to a partially specified target matrix in exploratory factor analysis. *Behavior Research Methods*, *47*, 494-505.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Osborne, J.W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research & Evaluation, 20*, 1-7.

Preacher, K.J., & MacCallum, R.C. (2003). Repairing Tom Swift's electric factor analysis machine. *Structural Equation Modeling*, *2*, 13-43.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reise, S.P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47,* 667–696.

Schweizer, K. (2012). On correlated errors. *European Journal of Psychological Assessment, 28,* 1-2.

Skrondal, A., & Laake, P. (2001). *Regression among factor scores. Psychometrika, 66,* 563-576.

Spearman, C. (1904). "General Intelligence" objectively determined and measured. *The American Journal of Psychology*, *15,* 201-292.

Thurstone, L.L. (1947). *Multiple Factor Analysis: A development and expansion of the vectors of the mind*. Chicago, IL: The University of Chicago.

Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles, CA: University of California.

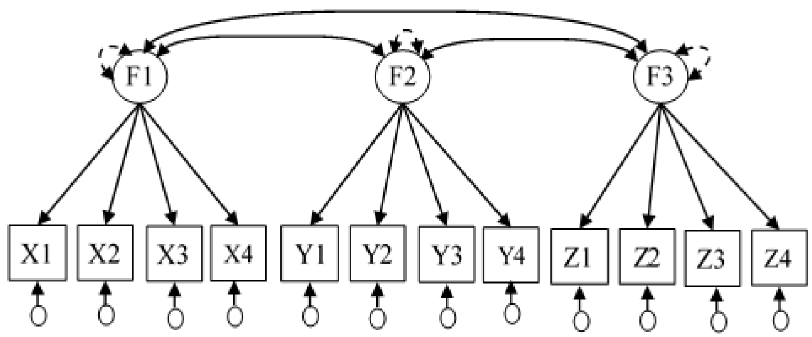

Figure 1a.
Confirmatory Factor Analytic (CFA) Measurement Model

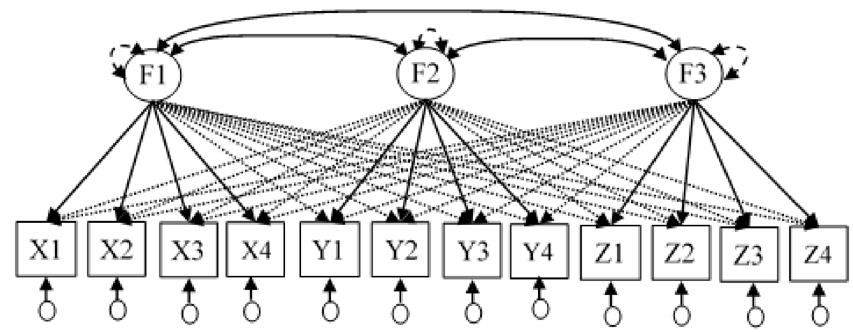

Figure 1b.
Exploratory Factor Analytic/Exploratory Structural Equation Modeling (EFA/ESEM) Measurement Model

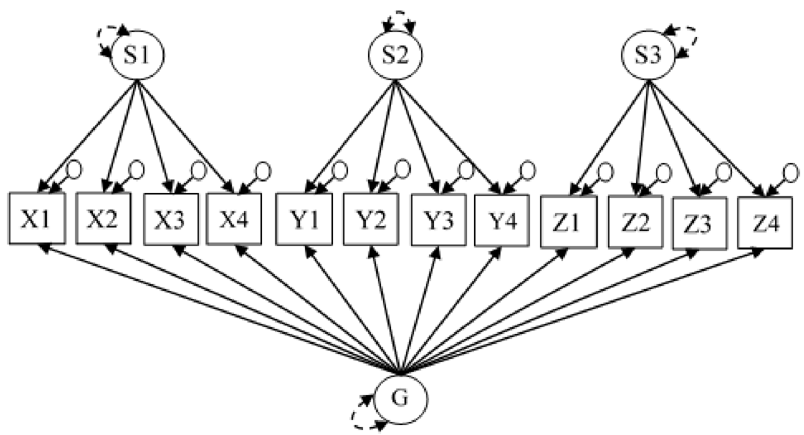

Figure 1c.
Bifactor Confirmatory Factor Analytic (Bifactor-CFA) Measurement Model

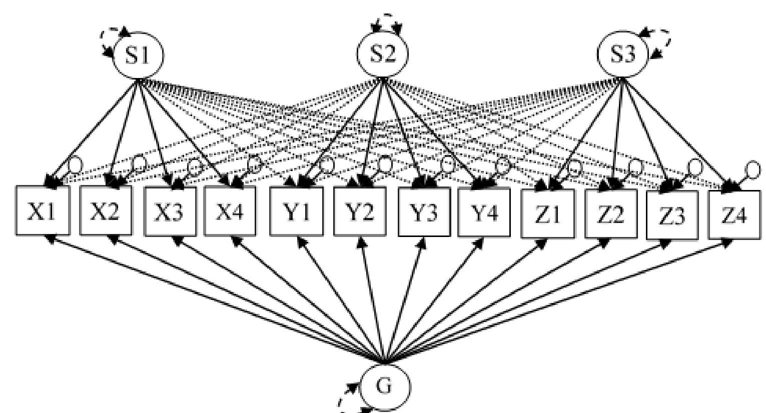

Figure 1d.
Bifactor Exploratory Structural Equation Modeling (Bifactor-ESEM) Measurement Model

Figure 1. *Alternative Measurement Models*
Note. X1-X4, Y1-Y4, Z1-Z4: Factor indicators; F1-F3: Factors; S1-S3: Specific factors (bifactor models); G: Global factors (bifactor models); Ovals: Latent factors; Squares: Observed variables; Full unidirectional arrows linking ovals and squares: Factor loadings; Dotted unidirectional arrows linking ovals and squares: Cross-loadings; Full unidirectional arrows linked to the items: Item uniquenesses; Bidirectional arrows linking the ovals: Factor covariances/correlations; Bidirectional arrows connecting a single oval: Factor variances.

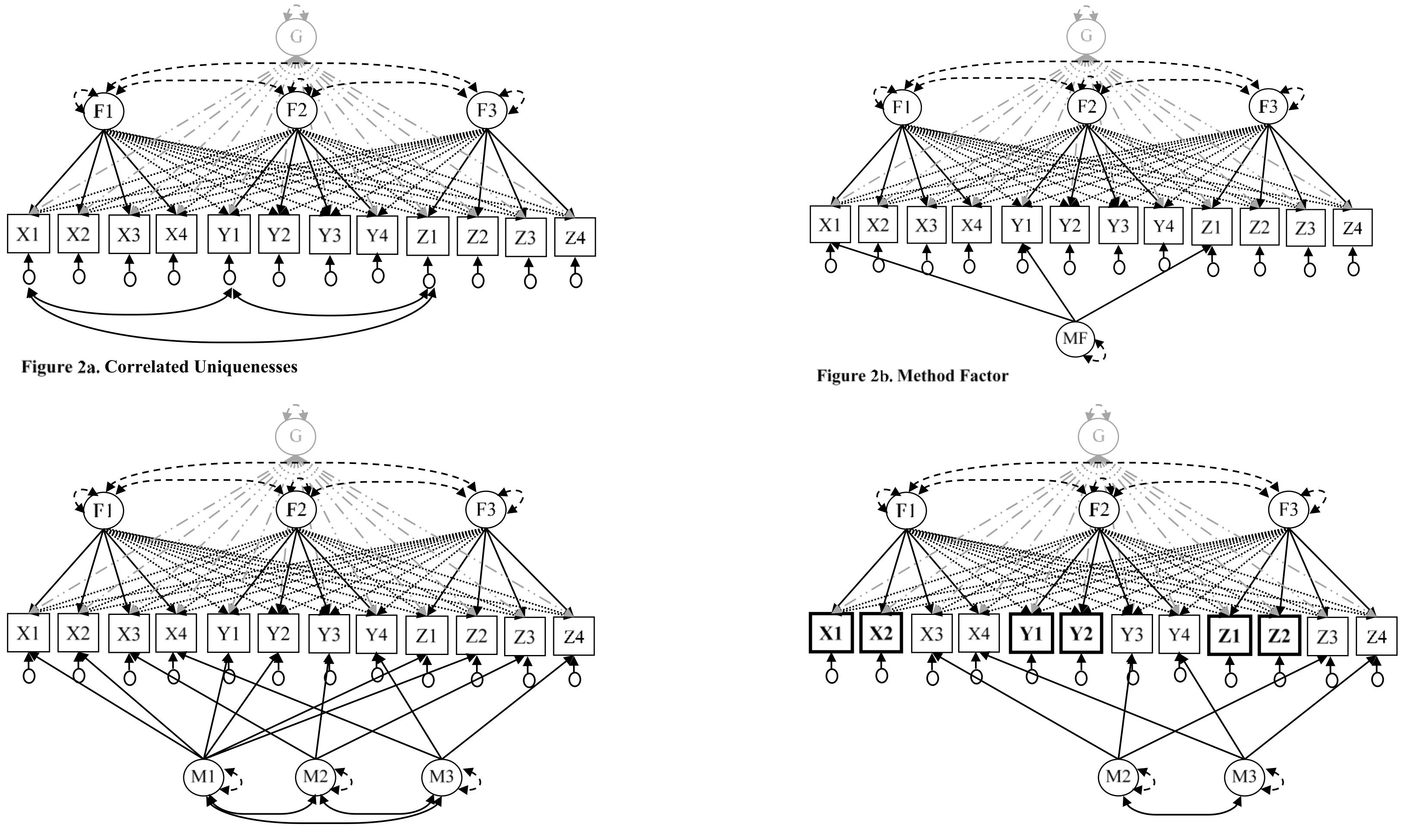

Figure 2a. Correlated Uniquenesses

Figure 2b. Method Factor

Figure 2c. Correlated Trait-Correlated Methods

Figure 2d. Correlated Trait-Correlated Methods Minus One

Figure 2. *Construct-Irrelevant Psychometric Multidimensionality*

Note. X1-X4, Y1-Y4, Z1-Z4: indicators; F1-F3: Factors (specific factors in bifactor models); G: Global factor (bifactor models); Ovals: Latent factors; Squares: Observed variables; Full unidirectional arrows linking ovals and squares: Loadings; Dotted unidirectional arrows linking ovals and squares: Cross-loadings (to be taken out in CFA or bifactor-CFA models); Full unidirectional arrows linked to the items: Uniquenesses; Bidirectional arrows linking the ovals: Covariances/correlations; Bidirectional arrows connecting a single oval: Variances; Dashed black arrows representing the factor correlations should be included in ESEM, but taken out in bifactor-ESEM; Greyscale component (including the dash-dot-dot arrows: G-factor loadings) should be included in bifactor models, but taken out otherwise; Bold squares reflect the anchoring indicators in the correlated traits correlated methods minus one model.

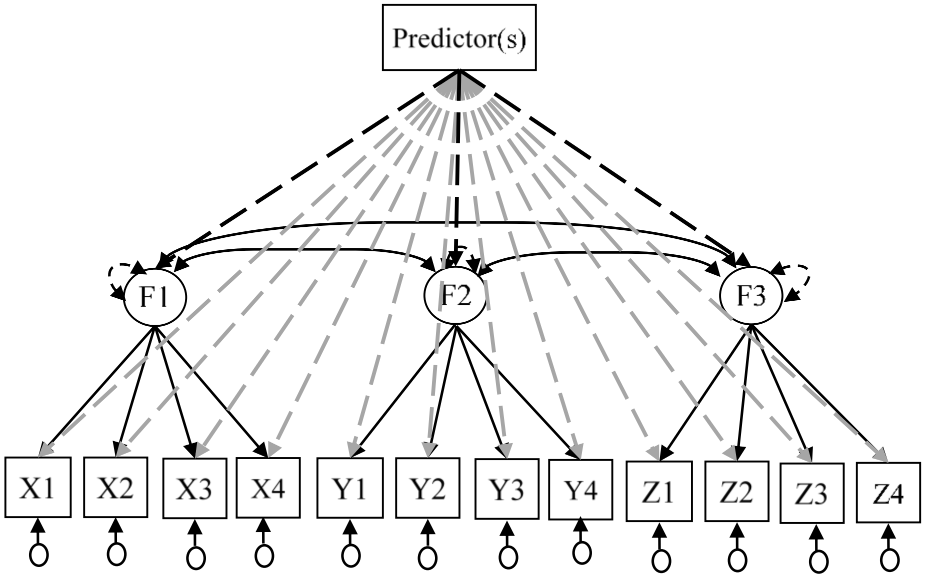

Figure 3. *Tests of Differential Item Functioning*
Note. X1-X4, Y1-Y4, Z1-Z4: Factor indicators; F1-F3: Factors (specific factors in bifactor models); Ovals: Latent factors; Squares: Observed variables; Full unidirectional arrows linking ovals and squares: Factor loadings; Full unidirectional arrows linked to the items: Uniquenesses; Bidirectional arrows linking the ovals: Factor covariances/correlations; Bidirectional arrows connecting a single oval: Factor variances; Dashed black arrows: Paths to be freely estimated in the invariant model (and fixed to zero otherwise).; Dashed greyscale arrows: Paths to be freely estimated in the saturated model (and fixed to zero otherwise); To simplify the figure, the cross-loadings and bifactor component were taken out, but can easily be added to the model estimation; For bifactor estimation, the factor correlations should be taken out, and predictive black arrows between the predictor and the global factor should be added.

Table 1
Goodness-of-Fit indices Associated with the Alternative Measurement Models

Description	χ^2 (*df*)	CFI	TLI	RMSEA	90% CI
Data Set #1 (ESEM population model)					
CFA	4212.946 (51)*	.898	.868	.090	.088; .093
ESEM	50.522 (33)	1.000	.999	.007	.003; .011
Bifactor-CFA	2709.330 (42)*	.935	.897	.080	.077; .082
Bifactor-ESEM	23.232 (24)*	1.000	1.000	.000	.000; .008
Data Set #2 (bifactor-ESEM population model)					
CFA	14084.303 (51)*	.849	.805	.166	.164; .168
ESEM	957.432 (33)*	.990	.980	.053	.050; .056
Bifactor-CFA	6942.682 (42)*	.926	.883	.128	.126; .131
Bifactor-ESEM	27.915 (24)	1.000	1.000	.004	.000; .009
Data Set #1 (ESEM Population Model): Measurement Invariance Across Groups					
Configural Invariance	76.025 (66)	1.000	1.000	.006	.000; .010
Weak Invariance	109.697 (93)	1.000	.999	.006	.000; .010
Strong Invariance	3013.353 (102)*	.930	.909	.076	.073; .078
Partial Strong Invariance	123.025 (101)	.999	.999	.007	.000; .010
Strict Invariance	135.819 (113)	.999	.999	.006	.000; .010
Latent Variances and Covariances Invariance	157.798(119)*	.999	.999	.008	.004; .011
Latent Means Invariance	159.461 (122)	.999	.999	.008	.004; .011
Data Set #2 (Bifactor-ESEM Population Model): Longitudinal Measurement Invariance					
Configural Invariance	178.426 (164)	1.000	1.000	.003	.000; .006
Weak Invariance	204.087 (196)	1.000	1.000	.002	.000; .005
Strong Invariance	3427.376 (204)*	.984	.978	.040	.039; .041
Partial Strong Invariance	217.187 (203)	1.000	1.000	.003	.000; .005
Strict Invariance	225.741 (215)	1.000	1.000	.002	.000; .005
Latent Variances and Covariances Invariance	240.907 (225)	1.000	1.000	.003	.000; .005
Latent Means Invariance	2699.858 (229)*	.987	.985	.033	.032; .034
Data Set #1 (ESEM Population Model): Tests of Differential Item Functioning					
Null Effects Model	3007.559 (45)*	.932	.883	.081	.079; .084
Saturated Model	46.329 (33)*	1.000	.999	.006	.000; .010
Invariant Model	2958.443 (42)*	.933	.876	.083	.081; .086
Partial Invariance Model	59.646 (41)	1.000	.999	.007	.002; .010

Note. * $p < .01$; CFA: Confirmatory factor analysis; ESEM: Exploratory structural equation modeling; χ^2: Robust chi-square test of exact fit; *df*: Degrees of freedom; CFI: Comparative fit index; TLI: Tucker-Lewis index; RMSEA: Root mean square error of approximation; 90% CI: 90% confidence interval.

Table 2

Standardized Parameter Estimates from with the Alternative Measurement Models

	CFA		Bifactor-CFA			ESEM				Bifactor-ESEM				
Items	λ	δ	G-λ	S-λ	δ	λ	λ	λ	δ	G-λ	S-λ	S-λ	S-λ	δ
Data set 1 (ESEM Population Model)														
X1	**.524**	.726	**.383**	**.377**	.711	**.490**	.221	-.106	.679	**.116**	**.491**	.258	-.030	.678
X2	**.629**	.604	**.325**	**.525**	.619	**.610**	-.094	.174	.581	**.321**	**.558**	-.051	.033	.582
X3	**.721**	.480	**.371**	**.599**	.504	**.698**	*.004*	.031	.501	**.242**	**.662**	.060	*.000*	.500
X4	**.871**	.242	**.384**	**.818**	.184	**.912**	-.044	-.034	.196	**.238**	**.862**	.035	-.045	.197
ω	.786			.727		.790					.772			
Y1	**.597**	.643	**.366**	**.463**	.651	.070	**.581**	-.045	.648	**.162**	.114	**.558**	-.017	.649
Y2	**.673**	.547	**.388**	**.538**	.560	-.185	**.691**	.153	.492	**.313**	-.139	**.625**	.022	.492
Y4	**.812**	.341	**.664**	**.495**	.314	.115	**.740**	.098	.350	**.366**	.156	**.700**	*.008*	.351
Y4	**.876**	.233	**.495**	**.768**	.166	.013	**.928**	-.099	.166	**.199**	.092	**.888**	-.018	.163
ω	.832			.752			.839					.823		
Z1	**.516**	.733	**.382**	**.382**	.709	.220	*.000*	**.478**	.680	**.531**	.166	-.021	**.081**	.683
Z2	**.516**	.734	**.225**	**.494**	.705	-.039	.021	**.548**	.702	**.405**	-.024	.013	**.820**	*.162*
Z3	**.598**	.642	**.105**	**.729**	.458	-.091	-.109	**.716**	.523	**.622**	-.154	-.164	**.153**	.540
Z4	**.885**	.217	**.518**	**.656**	.300	.026	.160	**.782**	.295	**.838**	-.040	.085	**.124**	.273
ω	.731		.783	.702				.743		.782			.456	
Data set 2 (Bifactor ESEM Population Model): Time 1														
X1	**.788**	.379	**.663**	**.504**	.306	**.615**	.389	-.030	.293	**.702**	**.438**	.156	-.064	.287
X2	**.719**	.484	**.323**	**.630**	.499	**.676**	-.153	.229	.437	**.436**	**.553**	-.103	.230	.440
X3	**.702**	.508	**.276**	**.657**	.492	**.763**	-.072	-.033	.481	**.353**	**.625**	.021	.071	.479
X4	**.890**	.209	**.414**	**.828**	.143	**.934**	-.032	-.017	.167	**.538**	**.740**	-.040	.038	.160
ω	.859			.826		.866					.803			
Y1	**.797**	.364	**.739**	**.298**	.364	.125	**.736**	-.017	.376	**.674**	.038	**.404**	-.091	.373
Y2	**.855**	.269	**.920**	***-.033***	.153	-.114	**.741**	.375	.173	**.859**	-.181	**.256**	.105	.152
Y4	**.811**	.342	**.734**	**.387**	.311	.086	**.776**	-.028	.354	**.590**	.056	**.635**	.007	.246
Y4	**.930**	.135	**.873**	**.334**	.125	.083	**.942**	-.064	.091	**.799**	-.021	**.476**	-.168	.107
ω	.912			.384			.911					.781		
Z1	**.791**	.374	**.383**	**.741**	.304	.197	-.175	**.788**	.307	**.432**	.242	*-.002*	**.729**	.223
Z2	**.784**	.386	**.513**	**.593**	.386	-.030	.084	**.758**	.386	**.567**	*-.012*	-.021	**.538**	.389
Z3	**.882**	.222	**.528**	**.738**	.176	-.034	-.025	**.952**	.142	**.648**	-.027	-.186	**.644**	.130
Z4	**.897**	.196	**.768**	**.509**	.151	.071	.331	**.684**	.159	**.788**	.042	.106	**.452**	.163
ω	.905		.925	.868				.911		.945			.861	

Note. CFA: Confirmatory factor analysis; ESEM: Exploratory structural equation modeling; WITH: Indicates a factor correlation; F1-F3: Factors 1 to 3 (defined as in Figure 1); G = Global factor estimated as part of a bifactor model; S = Specific factor estimated as part of a bifactor model; λ: Factor loading; δ: Item uniqueness; ω: Omega coefficient of model-based composite reliability; Main factor loadings are indicated in bold.