Online Conceptual Supplements for:

Chapter 27. Exploratory Structural Equation Modeling

Alexandre J.S. Morin, Substantive-Methodological Synergy Research Laboratory, Department of Psychology, Concordia University, Canada

Table of Contents

This is the prepublication version of the following document:

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

SECTION 1.

LIMITATIONS OF HIGHER-ORDER MODELS

The most frequently used approach to study the structure of hierarchically-ordered constructs relies on higher-order models, in which indicators are used to define a series of first-order factors, themselves used to estimate one (or more) second-order factor(s). Although never mentioned explicitly, first-order factors are always specified as orthogonal (i.e., uncorrelated with one another) in a higher-order model. Despite their intuitive appeal, higher-order models present two critical limitations.

First, in a higher-order model, the relation between any indicator and the second-order factor is indirect, and reflected by the product of the indicator's loading on the first-order factor (path a) and the loading of the first-order factor on the second-order factor (path b). Similarly, the relation between each indicator and the unique part of the first-order factor (i.e., unexplained by the second-order factor) is also an indirect effect, reflected by the product of the indicator's first-order factor loadings (path a) and the link between the first-order factor and its disturbance (path c). Because paths b and c are constant for all indicators of a single first-order factor, the ratio of variance explained by the second-order factor relative to that explained by the first-order factor is a constant for all indicators associated with a specific first-order factor (a*b/a*c can be reduced to the constant b/c). Even though this implicit proportionality constraint may hold in some rare circumstances and help to introduce parsimony, it is unlikely to be supported in many practical applications (Morin et al., 2016a; Reise, 2012) or to make sense logically or theoretically (Gignac, 2016).

Second, in higher-order models, the first-order factors reflect a combination of the variance explained by the second-order factor and of the variance uniquely explained by the first-order factor. To consider this second component on its own, one needs to consider the disturbance of the first-order factor, rather than the first-order factor itself. Because of this dual nature of the first-order factors, the joint inclusion of the first- and second-order factors in subsequent analyses creates a logical redundancy likely to result in flawed parameter estimates (Morin et al., 2016b, 2017).

As a result, higher-order models should only be used when they are clearly supported by theory or research. Even then, they should be contrasted with more flexible bifactor models before making a final decision.

References

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57-68.

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., Boudrias, J.-S., Marsh, H.W., Madore, I., & Desrumaux, P. (2016b). 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.

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

SECTION 2.
BIFACTOR MODELS, ORTHOGONALITY, VANISHING S-FACTORS, AND THE BIFACTOR S-1 FICTION

Over the years, various misconceptions have been vehiculated relative to bifactor models and are thus directly relevant to bifactor-ESEM models. These misconceptions are mainly related to the orthogonality assumption of bifactor models, relative to the need to estimate the factor correlations between the method factors in CTCM models, and to the mathematical possibility of relying on oblique forms of "bifactor" rotations (Jenrich & Bentler, 2011, 2012). Attempts to lump together all forms of models resulting in the disaggregation of true score variance into two components (global-specific *vs* trait-method) for the sake of simplicity has further contributed to muddling the relatively clear distinctions between these two types of models. Statistically, both types of models assume that the G-factor (in bifactor models) or the trait factor (in CTCM models) will be uncorrelated with the S-factors (in bifactor models) or the method factors (in CTCM models). The difference lies in the fact that bifactor models also set the S-factors to be uncorrelated with one another, while CTCM models allow the method factors to be correlated with one another. This statistical difference entails more important theoretical differences.

Substantively, bifactor models disaggregate multidimensional ratings into two distinct sets of theoretically meaningful constructs. The G-factor reflects the variance shared among all indicators, thus reflecting participants' global scores across all dimensions of the constructs (i.e., their global intelligence, self-concept, or motivation). The S-factors indicate the variance shared among a subset of indicators forming a subscale left unexplained by the G-factor. They reflect the specificity, or unique nature, of each subscale net of what it shares with the other subscales. Contrary to factors estimated in a non-bifactor model, which reflect the subscale-relevant variance in its entirety, the S-factors reflect the extent to which participants' scores on each dimension deviate from their scores across all dimensions (i.e., on the G-factor). These S-factors remain theoretically meaningful in their own right, showing the extent to which participants are uniquely characterized by each specific dimension (e.g., one might have a high level of IQ, and yet display a level of performance on memory tasks that is higher than expected given their global IQ). In contrast, CTCM models estimate meaningful trait factors from a series of indicators, while controlling for different sets of methodological artefacts (such as informant effects) which are not theoretically relevant themselves.

Confusions between both types of models have always been present in research, as illustrated by studies including correlated "specifics" to study meaningful global and specific constructs (e.g., Brunner et al., 2009; 2010). However, this issue has recently been further muddled by the suggestion to rely on bifactors (S-1) models (removing one of the S-factors, but adding correlations among the remaining S-factors), similar to CTC(M-1) models (e.g., Burns et al., 2020; Eid et al., 2017). This proposition has emerged from the observation of bifactor results erroneously described as "anomalous" (e.g., Eid et al., 2017; Markon, 2019). Three categories of so-called "anomalous" results were identified: (a) "vanishing" S-factors (i.e., models in which one of the S-factor is weakly defined by its indicators); (b) "irregular" factor loading patterns that do not match one's expectations or the loadings observed in non-bifactor correlated-factors models; (c) the observation of correlations between the specific factors.

Clearly, issue *c* above can only stem from researchers' reliance on improper bifactor models in which the S-factors were allowed to be correlated in the first place, which should not have happened. Likewise, it is hard to see how issue *b* can be a problem, at least statistically speaking. Although data frequently fails to match hypotheses, why would one expect S-factor loadings to be identical to their correlated-factors counterpart? When moving from a correlated-factors model to a bifactor model, one explicitly seeks to change the meaning of the S-factors as reflecting the unique nature of the subscale disaggregated from the global component. In this context, we should expect some indicators to retain more (showing stronger S-factor loadings) or less (showing stronger G-factor loadings and weak S-factor loadings) specificity. Otherwise, we should not pursue a bifactor representation of the data. Lacking prior evidence, it is typically difficult to develop clear a priori hypotheses in this regard. However, moving from one type of model to the other, it seems logical to expect some differences.

For the same reasons, the observation of vanishing S-factors (issue *a*) is also to be expected in bifactor research, and only shows that limited specificity remains at the level of the S-factor once the G-factor has been taken into account. A "vanishing" S-factor indicates that the items associated with this specific dimension mainly contribute to define the G-factor and retain very little specificity beyond that. Consider the example of the satisfaction of the needs for competence, relatedness, and autonomy, which have been repeatedly shown to possess a bifactor structure where the global factor reflects global levels of need satisfaction across all needs, and the S-factors reflect deviations in the satisfaction of each specific need relative to that global level. This area of research typically results in the identification of one "vanishing" S-factor, so that this phenomenon can now be expected *a priori*. However, this "vanishing" S-factor depends on the nature of the sample. Thus, among generic populations of workers and university students, the need for autonomy tends to present a strong alignment with global levels of need satisfaction (resulting in a vanishing S-factor) (Gillet et al., 2019; 2020; Sánchez-Oliva et al., 2017), whereas the same happens to the need for relatedness among younger populations of students (Garn et al., 2019) and nurses (Huyghebaert-Zouaghi et al., 2021).

To explain these "anomalies," Eid et al. (2017) transposed stochastic measurement theory, which assumes that the indicators of a reflective construct should be conceptualized as a random sample of all possible indicators of that construct (aligned with CTT), to suggest that the various subdomains (represented by the S-factors) incorporated in bifactor models should be seen as a random sample of all possible subdomains of that construct. To address the lack of subdomain "interchangeability," they propose removing one of the S-factors, allowing the remaining S-factors to correlate, thereby "anchoring" the meaning of the G-factor into the omitted domain, as in the CTC(M-1) domain.

Unfortunately, although this assumption matches CTCM models in which the various "methods" or "raters" are seen as a random sample of all possible "methods" or "raters," it is flawed in relation to bifactor models. In a bifactor model, the whole set of indicators are assumed to be a random sample of all possible indicators of the G-factor. Likewise, each subset of indicators is assumed to be a random sample of all possible indicators of each S-factor. However, this assumption does not extend to the domains, dimensions, or S-factors covered in the measure. First, in a bifactor model, all factors are directly estimated at the indicator level (contrary to the higher-order model in which they are estimated at the factor level, and for which this assumption of "interchangeability" might be more relevant). Second, in a bifactor model, one typically assumes that all relevant subdomains are covered, at least from the perspective of the measure used. For instance, in the need satisfaction example used before, Self-Determination Theory (Ryan & Deci, 2017) assumes that these three needs form a complete set. Arguably, one will not necessarily be able to explicitly focus on all possible subdomains. However, even then, bifactor applications will usually assume that the main domains are reasonably covered to be able to interpret the G-factor as reflecting scores obtained across all domains. As a result, in a bifactor model, the G-factor does not need an anchor to be interpretable.

Furthermore, due to the orthogonality of a bifactor model, the definition of the G-factor and of the remaining S-factors remains unchanged following the removal of one of the S-factor. More precisely, in bifactor models, the clean partitioning of the variance explained by the global and specific constructs is made possible by the orthogonality of the factors, as this orthogonality forces the covariance shared among all items to be fully absorbed into the G-factor (resulting in a clear definition of the G-factor as reflecting the variance shared among all of its subdomains), while the S-factors represent the covariance shared among a subset of items but not with the other subsets. As such, taking out one of the S-factors (such as a "vanishing" S-factor) will not change the meaning of the G-factor as reflecting what is shared among *all* indicators. This stability has been demonstrated by Arens and Morin (2017) and Morin et al. (2020).

In contrast, CTCM models "force" construct-relevant variance to be absorbed by the trait factor and rely on the method factors to control for inter-rater or inter-method differences. However, by allowing these method factors to be correlated, the CTCM adds a third source of covariance, which makes it difficult to properly interpret the trait factor (and to converge). The trait factor should be used to estimate what is shared among all indicators, but how to achieve this when all items are already allowed to share something with all other items (by allowing the method factors to correlate) beyond the trait factor is unclear. The CTC(M-1) model provides a solution. By anchoring the definition of the trait factor in one of the "methods," this model clarifies the meaning of the trait (i.e., reflecting the ratings obtained using the anchoring method or informant, and what they share with the other types of methods or informants) and method (i.e., reflecting the extent to which the ratings provided by each type of informant or method deviate from the ratings provided by the anchoring method or informant) factors. This phenomenon is related to the non-orthogonality of the method factors, which "push" into the trait factor the information that is shared among all indicators as well as the information that is shared among the indicators of the referent method. Using real data, Morin et al. (2020) demonstrated this key difference, showing that the anchoring effect only occurred when correlations were included among the S-factors/method factors, but not in the absence of these correlations. Importantly, moving from an orthogonal bifactor model to a so-called "bifactor (S-1)" model in which the S-factors are allowed to correlate is likely to change the meaning of the G-factor in a way that no longer makes theoretical sense. Why would we want to "anchor" IQ scores ratings into a specific subtest (e.g., memory)?

Further Demonstration

We further illustrate these differences using three data sets simulated according to a (1) bifactor CFA model (relying on parameter estimates corresponding to that of Data 2 but without the cross-loadings), a (2) CTCM model (relying on the same parameter estimates) in which the method factors share correlation of .35, and (3) another CTCM model relying on the same parameter estimates but correlations of .50 among the method factors. We also simulated three additional data sets almost identical to the previous ones, but in which the S-factor or method-factor to be removed was a "vanishing" one (with loadings varying from .100 to .200). The results (presented in Tables S1 and S2 of these Conceptual Supplements) are pretty straightforward and show that: (a) no "anchoring" effects occur when the omitted S-factor or method factor is a "vanishing" factor; (b) for data simulated according to models including three strong S-factors or method factors, the removal of one of those factors results in a substantial decrease in model fit; (c) when the data is simulated according to a bifactor model with three strong S-factors, taking out one S-factors results in some minor changes in the size of the factor loadings that do not modify the meaning of the factors, whereas these changes are more pronounced when the remaining S-factors are allowed to be correlated; (d) for data simulated according to a CTCM model with three strong method factors, the removal of one method factor results in a clear "anchoring" effect but only when the remaining method factors are allowed to be correlated, not when they are orthogonal; (e) removing a "vanishing" factor does not modify the remaining factor loadings in any way. These results thus support the previous discussion.

References

Arens, A.K., & Morin, A.J.S. (2017). Improved representation of the Self-Perception Profile for Children through bifactor exploratory structural equation modeling. *American Educational Research Journal*, *54*, 59-87.

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology*, *102*, 964-981.

Brunner, M., Keller, U., Hornung, C., Reichert, M., & Martin, R. (2009). The cross-cultural generalizability of a new structural model of academic self-concepts. *Learning and Individual Differences*, *19*, 387-403.

Burns, G.L., Geiser, C., Servera, M., Becker, S.P., & Beauchaine, T.P. (2020). Application of the bifactor S-1 model to multisource ratings of ADHD/ODD symptoms. *Journal of Abnormal Child Psychology*, *48*, 881-894.

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*, 541-562.

Garn, A.C., Morin, A.J.S., & Lonsdale, C. (2019). Basic psychological need satisfaction toward learning: A longitudinal test of mediation using bifactor exploratory structural equation modeling. *Journal of Educational Psychology*, *111*, 354-372.

Gillet, N., Morin, A.J.S., Choisay, F., & Fouquereau, E. (2019). A person-centered representation of basic need satisfaction balance at work. *Journal of Personnel Psychology*, *18*, 113-128.

Gillet, N., Morin, A.J.S., Huyghebaert-Zouaghi, T., Alibran, E., Barrault, S., & Vanhove-Meriaux, C. (2020b). Students' need satisfaction profiles: Similarity and change over the course of a university semester. *Applied Psychology*, *69*, 1396-1437.

Huyghebaert-Zouaghi, T., Morin, A.J.S., Forest, J., Fouquereau, E., & Gillet, N. (2021, in press). A longitudinal examination of nurses' need satisfaction profiles: A latent transition analysis. *Current Psychology*. doi: 10.1007/s12144-020-00972-1.

Jennrich, R.I., & Bentler, P.M. (2011). Exploratory bi-factor analysis. *Psychometrika*, *76*, 537–549.

Jennrich, R.I., & Bentler, P.M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika*, *77*, 442–454.

Markon, K.E. (2019). Bifactor and hierarchical models: Specification, inference, and interpretations. *Annual Review of Clinical Psychology*, *15*, 51-69.

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 Ed., Vol. 2* (pp. 1044-1073). Wiley.

Ryan, R.M., & Deci, E.L. (2017). *Self-determination theory. Basic psychological needs in motivation, development, and wellness*. New York: Guildford.

Sánchez-Oliva, D., Morin, A.J.S., Teixeira, P.J., Carraça, E.V., Palmeira, A.L., & Silva, M.N. (2017). A bifactor exploratory structural equation modeling representation of the structure of the basic psychological needs at work scale. *Journal of Vocational Behavior*, *98*, 173-187.

Table S1.

Additional Simulated Data Results: Model fit indices

Description	χ^2 (*df*)	CFI	TLI	RMSEA	90% CI
Data Simulated: Bifactor Model, Strong S-Factors					
Bifactor	197.849 (42)*	.998	.997	.019	.017; .022
Bifactor (S-1)	2667.066 (46)*	.970	.957	.075	.073; .078
CTCM	61.933 (39)	1.000	1.000	.008	.004; .011
CTC(M-1)	1986.665 (45)*	.978	.967	.066	.063; .068
Data Simulated: Bifactor Model, One Vanishing S-Factor					
Bifactor	58.880 (42)	1.000	1.000	.006	.001; .010
Bifactor (S-1)	93.423 (46)*	.999	.999	.010	.007; .013
CTCM	72.118 (39)*	1.000	.999	.009	.006; .013
CTC(M-1)	91.029 (45)*	.999	.999	.010	.007; .013
Data Simulated: CTCM, r = .35, Strong Method Factors					
Bifactor	444.755 (42)*	.995	.993	.031	.028; .034
Bifactor (S-1)	2525.084 (46)*	.972	.960	.073	.071; .076
CTCM	64.333 (39)*	1.000	1.000	.008	.004; .011
CTC(M-1)	1806.940 (45)*	.980	.971	.063	.060; .065
Data Simulated: CTCM, r = .50, Strong Method Factors					
Bifactor	770.356 (42)*	.992	.987	.042	.039; .044
Bifactor (S-1)	2413.753 (46)*	.974	.963	.072	.069; .074
CTCM	66.414 (39)*	1.000	.999	.008	.005; .012
CTC(M-1)	1711.283 (45)*	.982	.973	.061	.058; .063
Data Simulated: CTCM, r = .35, One Vanishing Method Factor					
Bifactor	301.042 (42)*	.996	.994	.025	.022; .028
Bifactor (S-1)	390.280 (46)*	.995	.993	.027	.025; .030
CTCM	61.173 (39)	1.000	.999	.008	.004; .011
CTC(M-1)	77.518 (45)*	1.000	.999	.009	.005; .012
Data Simulated: CTCM, r = .5, One Vanishing Method Factor					
Bifactor	584.291 (42)*	.992	.988	.036	.033; .039
Bifactor (S-1)	748.543 (46)*	.990	.986	.039	.037; .042
CTCM	56.747 (39)	1.000	1.000	.007	.002; .010
CTC(M-1)	93.130 (45)*	.999	.999	.010	.007; .013

Note. * $p < .01$; Bifactor (S-1): Bifactor model minus one specific factor; CTCM: Correlated trait correlated methods model; CTC(M-1): CTCM model minus one method factor; S-factor: Specific factor from a bifactor model; 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 S2.

Additional Simulated Data Results: Parameter estimates (1/3)

	Bifactor		Bifactor (S-1)				CTCM		CTC(M-1)			
Indicators	GF λ	SF λ	GF λ	SF λ	ΔGF λ	ΔSF λ	TF λ	MF λ	TF λ	MF λ	ΔTF λ	ΔMF λ
Data Simulated: Bifactor Model, Strong S-Factors												
X1	.719	.348	.618	.468	-.101	.120	.694	.403	.603	.495	-.091	.092
X2	.535	.469	.473	.530	-.062	.061	.494	.511	.459	.545	-.035	.034
X3	.442	.568	.393	.596	-.049	.028	.390	.606	.379	.605	-.011	-.001
X4	.647	.652	.573	.721	-.074	.069	.588	.706	.556	.727	-.032	.021
Y1	.634	.453	.544	.558	-.090	.105	.598	.499	.531	.571	-.067	.072
Y2	.820	.345	.699	.521	-.121	.176	.806	.393	.684	.541	-.122	.148
Y3	.563	.625	.491	.658	-.072	.033	.502	.687	.478	.665	-.024	-.022
Y4	.747	.546	.641	.677	-.106	.131	.706	.593	.625	.689	-.081	.096
Z1	.571	.622	.791		.220		.504	.686	.797		.293	
Z2	.626	.458	.772		.146		.584	.509	.774		.190	
Z3	.749	.546	.915		.166		.699	.606	.922		.223	
Z4	.816	.340	.881		.065		.792	.403	.876		.084	
Data Simulated: Bifactor Model, One Vanishing S-Factor												
X1	.695	.401	.689	.408	-.006	.007	.696	.398	.687	.411	-.009	.013
X2	.496	.510	.493	.513	-.003	.003	.499	.507	.491	.515	-.008	.008
X3	.390	.606	.387	.608	-.003	.002	.396	.602	.385	.609	-.011	.007
X4	.589	.705	.584	.709	-.005	.004	.594	.701	.582	.711	-.012	.010
Y1	.595	.504	.587	.513	-.008	.009	.614	.480	.585	.515	-.029	.035
Y2	.800	.403	.791	.418	-.009	.015	.814	.373	.789	.421	-.025	.048
Y3	.496	.691	.489	.693	-.007	.002	.523	.670	.487	.695	-.036	.025
Y4	.700	.602	.691	.613	-.009	.011	.722	.575	.689	.615	-.033	.040
Z1	.517	.045	.520		.003		.514	.062	.520		.006	
Z2	.586	.148	.600		.014		.589	.153	.600		.011	
Z3	.692	.273	.711		.019		.700	.187	.712		.012	
Z4	.795	.110	.806		.011		.794	.132	.807		.013	

Note. Bifactor (S-1): Bifactor model minus one specific factor; CTCM: Correlated trait correlated methods model; CTC(M-1): CTCM model minus one method factor; S-factor: Specific factor from a bifactor model; GF: Global factor from a bifactor model; SF: Specific factor from a bifactor model; TF: Trait factor from a CTCM model; MF: Method factor from a CTCM model; *r*: correlation; λ: Factor loading; Δ: Change between the basic model (bifactor or CTCM) and the reduced models (bifactor (S-1) or CTC(M-1)) in the size of the factor loadings.

Table S2. (cont.)

Additional Simulated Data Results: Parameter estimates (2/3)

	Bifactor		Bifactor (S-1)				CTCM		CTC(M-1)			
Indicators	GF λ	SF λ	GF λ	SF λ	ΔGF λ	ΔSF λ	TF λ	MF λ	TF λ	MF λ	ΔTF λ	ΔMF λ
Data Simulated: CTCM, r = .35, Strong Method Factors												
X1	.736	.300	.653	.408	-.083	.108	.693	.406	.640	.442	-.053	.036
X2	.567	.429	.517	.481	-.050	.052	.493	.512	.505	.502	.012	-.010
X3	.486	.530	.452	.556	-.034	.026	.389	.606	.435	.566	.046	-.040
X4	.695	.602	.635	.659	-.060	.057	.585	.709	.620	.673	.035	-.036
Y1	.663	.409	.594	.508	-.069	.099	.596	.501	.577	.523	-.019	.022
Y2	.832	.300	.738	.464	-.094	.164	.806	.394	.721	.487	-.085	.093
Y3	.613	.563	.552	.593	-.061	.030	.498	.690	.542	.610	.044	-.080
Y4	.781	.500	.700	.622	-.081	.122	.704	.594	.680	.636	-.024	.042
Z1	.621	.561	.786		.165		.501	.690	.798		.297	
Z2	.655	.414	.768		.113		.583	.509	.773		.190	
Z3	.784	.498	.907		.123		.697	.607	.922		.225	
Z4	.829	.293	.871		.042		.792	.404	.876		.084	
Data Simulated: CTCM, r = .50, Strong Method Factors												
X1	.753	.244	.694	.340	-.059	.096	.691	.409	.678	.380	-.013	-.029
X2	.600	.381	.568	.427	-.032	.046	.491	.513	.552	.451	.061	-.062
X3	.532	.482	.508	.501	-.024	.019	.388	.605	.491	.517	.103	-.088
X4	.744	.541	.706	.593	-.038	.052	.583	.712	.685	.608	.102	-.104
Y1	.691	.357	.638	.443	-.053	.086	.594	.503	.624	.466	.030	-.037
Y2	.841	.252	.771	.394	-.070	.142	.804	.396	.756	.423	-.048	.027
Y3	.665	.486	.624	.523	-.041	.037	.494	.695	.607	.543	.113	-.152
Y4	.815	.447	.754	.551	-.061	.104	.703	.594	.736	.571	.033	-.023
Z1	.673	.486	.795		.122		.498	.695	.801		.303	
Z2	.684	.363	.769		.085		.583	.508	.773		.190	
Z3	.819	.443	.914		.095		.696	.607	.922		.226	
Z4	.840	.242	.877		.037		.791	.406	.875		.084	

Note. * $p < .01$; Bifactor (S-1): Bifactor model minus one specific factor; CTCM: Correlated trait correlated methods model; CTC(M-1): CTCM model minus one method factor; S-factor: Specific factor from a bifactor model; GF: Global factor from a bifactor model; SF: Specific factor from a bifactor model; TF: Trait factor from a CTCM model; MF: Method factor from a CTCM model; r: correlation; λ: Factor loading; Δ: Change between the basic model (bifactor or CTCM) and the reduced models (bifactor (S-1) or CTC(M-1)) in the size of the factor loadings.

Table S2. (cont.)

Additional Simulated Data Results: Parameter estimates (3/3)

	Bifactor		Bifactor (S-1)				CTCM		CTC(M-1)			
Indicators	GF λ	SF λ	GF λ	SF λ	ΔGF λ	ΔSF λ	TF λ	MF λ	TF λ	MF λ	ΔTF λ	ΔMF λ
Data Simulated: CTCM, r = .35, One Vanishing Method Factor												
X1	.734	.319	.728	.331	-.006	.012	.694	.399	.709	.370	.015	-.029
X2	.552	.449	.546	.456	-.006	.007	.500	.505	.522	.482	.022	-.023
X3	.461	.553	.454	.559	-.007	.006	.397	.601	.424	.582	.027	-.019
X4	.669	.629	.661	.639	-.008	.010	.595	.700	.626	.673	.031	-.027
Y1	.657	.418	.642	.441	-.015	.023	.622	.469	.618	.474	-.004	.005
Y2	.845	.292	.830	.325	-.015	.033	.823	.353	.813	.366	-.010	.013
Y3	.587	.610	.570	.621	-.017	.011	.536	.659	.536	.654	.000	-.005
Y4	.774	.503	.757	.530	-.017	.027	.735	.557	.730	.565	-.005	.008
Z1	.504	.119	.517		.013		.513	.092	.521		.008	
Z2	.576	.190	.597		.021		.592	.112	.603		.011	
Z3	.686	.243	.711		.025		.704	.150	.718		.014	
Z4	.773	.202	.794		.021		.792	.146	.806		.014	
Data Simulated: CTCM, r = .5, One Vanishing Method Factor												
X1	.751	.272	.746	.285	-.005	.013	.695	.398	.720	.348	.025	-.050
X2	.579	.412	.573	.421	-.006	.009	.501	.504	.537	.465	.036	-.039
X3	.499	.519	.489	.529	-.010	.010	.397	.601	.442	.568	.045	-.033
X4	.710	.583	.699	.596	-.011	.013	.596	.700	.647	.653	.051	-.047
Y1	.687	.365	.669	.397	-.018	.032	.616	.476	.634	.452	.018	-.024
Y2	.865	.224	.849	.270	-.016	.046	.820	.361	.825	.336	.005	-.025
Y3	.634	.557	.611	.578	-.023	.021	.527	.668	.558	.638	.031	-.030
Y4	.811	.441	.790	.480	-.021	.039	.729	.564	.749	.538	.020	-.026
Z1	.496	.151	.514		.018		.510	.107	.521		.011	
Z2	.570	.206	.594		.024		.592	.119	.604		.012	
Z3	.681	.252	.708		.027		.701	.170	.721		.020	
Z4	.760	.246	.786		.026		.792	.144	.805		.013	

Note. * $p < .01$; Bifactor (S-1): Bifactor model minus one specific factor; CTCM: Correlated trait correlated methods model; CTC(M-1): CTCM model minus one method factor; S-factor: Specific factor from a bifactor model; GF: Global factor from a bifactor model; SF: Specific factor from a bifactor model; TF: Trait factor from a CTCM model; MF: Method factor from a CTCM model; r: correlation; λ: Factor loading; Δ: Change between the basic model (bifactor or CTCM) and the reduced models (bifactor (S-1) or CTC(M-1)) in the size of the factor loadings.

Section 3.
Reliability Estimation in EFA/ESEM and Bifactor-ESEM

Reliability is a core component of psychometric investigations (Raykov, Chapter 25, this volume). Unfortunately, Cronbach alpha remains the most reported reliability indicator, despite its multiple limitations (Sijtsma, 2009). Without going into all of these limitations, alpha first assumes that all indicators are equivalent (i.e., equal factor loadings) and linked to a single dimension (i.e., unidimensional). Anyone familiar with factor analyses should quickly grasp the unrealism of the first assumption. The present chapter should make it easy to understand that the second assumption is unrealistic for measures following a bifactor, EFA/ESEM, or bifactor-ESEM structure. Among alternatives, McDonald (1970) omega (ω) coefficient seems to be the most flexible, in addition to being easy to calculate and directly connected to the properties of the retained measurement model (Dunn et al., 2014). ω is calculated using the factor loadings (λ_i) and uniquenesses (δ_{ii}): $\omega = (\Sigma|\lambda_i|)^2 / ([\Sigma|\lambda_i|]^2 + \Sigma\delta_{ii})$. As a result, ω is directly connected to CTT definition of reliability presented earlier ($r_{xx} = \sigma^2_{true}/\sigma^2_{total} = \omega$) if we assume that $(\Sigma|\lambda_i|)^2 = \sigma^2_{true}$ and $([\Sigma|\lambda_i|]^2 + \Sigma\delta_{ii}) = \sigma^2_{total}$ at the scale level, and that $\lambda_i^2 = \sigma^2_{true}$ and δ_i = at σ^2_{error} the item level. In bifactor models, both the G- and the S- factors are thus assumed to represent σ^2_{true}. To account for this duality, some have proposed alternative specifications of ω (e.g., Reise et al., 2013; Rodriguez et al., 2016). The first is related to the G-factor for models including *q* S-factors:

$$\omega_h = \frac{\left(\sum|\lambda_{gi}|\right)^2}{\left(\sum|\lambda_{gi}|\right)^2 + \left(\sum|\lambda_{s1i}|\right)^2 + \left(\sum|\lambda_{s2i}|\right)^2 + \cdots + \left(\sum|\lambda_{sqi}|\right)^2 + \left(\sum \delta_{ii}\right)}$$

The second is related to the S-factors:

$$\omega_s = \frac{\left(\sum|\lambda_{s1i}|\right)^2}{\left(\sum|\lambda_{gi}|\right)^2 + \left(\sum|\lambda_{s1i}|\right)^2 + \left(\sum \delta_{ii}\right)}$$

Despite their intuitive appeal, these alternative coefficients are irremediably flawed in failing to match one critical corollary of CTT definition of reliability: $r_{xx} = \sigma^2_{true}/\sigma^2_{total}$ also means that $1 - r_{xx} = \sigma^2_{error}$ (or $1 - \omega = \sigma^2_{error}$) (Morin et al., 2020). Because the denominator of ω_h and ω_s includes components of σ^2_{true} related, respectively, to the S- and G- factors, this corollary is no longer true when ω_h and ω_s are considered. As a result, these coefficients will necessarily underestimate the reliability of the factors unless the converse source of σ^2_{true} (i.e., due to the G-factor in the calculation of ω_s and to the S-factor in the calculation of ω_h) is equal to zero, and will reduce linearly as this converse source of σ^2_{true} increases. In contrast, the classical ω provides a far more direct way to estimate the extent to which scores on a G- or S- factors reflect σ^2_{true} relative to σ^2_{error}, without being contaminated by the fact that σ^2_{true} is distributed across more than one factor. As a result, we argue that ω should always be reported. As long as ω is reported, authors should be free (but not obligated) to report ω_h and ω_s, as long as they properly interpret these coefficients as reflecting the comparative strength of each type of factor at explaining σ^2_{true}, rather than as indicators of reliability.

When moving to the cross-loading component of EFA/ESEM or bifactor-ESEM, which also arguably reflects some form of σ^2_{true}, we are unfortunately not aware of any recommendations regarding how these cross-loadings should be taken into account in the calculation of ω. Clearly, although a cross-loading indicates that an indicator shares some construct-relevant association with a secondary factor, this association remains extraneous to the main definition of the constructs themselves (i.e., it would not need to be incorporated into the model if the secondary construct was not also included in the model). As such, they do not reflect properties of scores on the construct for which ω is calculated, nor do they reflect random measurement error. Rather, they are simply incorporated into the model to control for the fallible nature of indicators. In practice, we currently recommend ignoring these cross-loadings in the calculation of ω (Morin et al., 2020), but note that this is a matter of common sense and an issue on which additional statistical research is needed.

References

Dunn, T.J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105,* 399-412.

Sijtsma, K. (2009). On the use, misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74,* 107-120.

McDonald, R.P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis,

and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*, 1-21.
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 Ed.*, *Vol. 2* (pp. 1044-1073). Wiley.
Reise, S.P., Bonifay, W.E., & Haviland, M.G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*, 129-140.
Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*, 137-150.

Section 4.
Power Analyses in EFA/ESEM and Bifactor-ESEM

Power Analyses

Power (Hancock & Feng, Chapter 9, this volume) is to the ability to detect as statistically significant effects present in the population. Power is known to be impacted by a variety of elements including sample size, effects size, the number of indicators per factor, the number of factors, the strength of the factor loadings, the quantity and type of missing data (Enders, Chapter 12, this Volume), and so on. Power analyses can be used to determine sample size requirements when planning a study (*a priori*), or to assess the power linked to specific aspects of the analyses once the data has been collected (*post hoc*). Power can be assessed in relation to model fit or model parameters, using three main tools for which introductory presentations have been provided by Myers et al. (2011, 2016, 2018): (a) Tables; (b) online calculators; (c) Monte-Carlo simulations (Leite, Bandalos & Shen, Chapter 6, this volume).

Elsewhere (Morin et al., 2020) we provided a quick review of these various methods, but timidly highlighted that power analyses might not be relevant for all types of research involving EFA/ESEM or bifactor-ESEM. We have since come to observe that requests to conduct *post hoc* power analyses are becoming more frequent from reviewers and editors alike. We thus take a stronger stand in this chapter to highlight the unrealistic and unnecessary nature of this requirement.

First, as noted above, power analyses depend on a great variety of elements that are typically impossible to know beforehand for *a priori* power analyses and very hard to account for in *post hoc* power analyses. Second, statistical simulation studies have shown that EFA-based measurement tends to be very robust to very small sample sizes (e.g., de Winter et al., 2009). Third, power is about the ability to detect as statistically significant effects present in the population model. Thus, a lack of power does not result in biased parameter estimates, but rather in inflated standard errors, leading to non-statistically significant results. As such, the estimation of statistically significant relations, coupled with the observation that standard errors are not larger than they should be, provides a sufficient proof that power is not an issue. Fourth, in EFA/ESEM and bifactor-ESEM measurement models, the interpretation of the results does not rely on statistical significance but on the relative size of the loadings, cross-loadings, and factor correlations. As such, power is a non-issue for measurement models' comparisons relying on the sequential strategy outlined above and may (but see our next point) only become an issue when moving to more complex predictive models.

Fifth, power (or lack thereof) is only one problem that might emerge from small sample sizes. In practice, a far more severe type of problem tends to occur when trying to do too much with a sample size that is too small: Nonconvergence (e.g., Chen et al., 2001). There is always a limit to the type of model that can be estimated using any specific sample. When one goes beyond this limit, analyses simply stop converging and no attempt to help the model achieve convergence (iterations, convergence criteria, constraints) will solve this problem. We have frequently faced this problem with all types of sample sizes (even with very large samples). When this happens, the solutions are to simplify the model, adopt a more parsimonious approach, divide the larger model into sub-models, and/or revise expectations. The key issue is that convergence problems typically arise well before power becomes an issue. As a result, the ability to produce research that includes meaningful, and statistically significant, results based on converging solutions is, in and of itself, an important safeguard against a lack of power.

When then, are power analyses useful? First, when planning a data collection process (or preparing a funding application) to ensure the recruitment of a sample large enough to detect the effects of interest. For costly data collections involving hard to access populations, *a priori* power analyses are a good way to save money and maximize returns on investments, especially when one is interested in the detection of relatively small effects. Second, when one wants to publish null findings (i.e., to demonstrate that a relation expected by theory does not occur) using sample sizes that are not plainly and undoubtedly "large enough," one must demonstrate that power is not the cause of this lack of statistical significance. Beyond this, researchers experiencing severe convergence difficulties, observing large standard errors, and obtaining moderately large, or very large, coefficients that are flagged as non-statistically significant should consider simplifying their models, rather than estimating power. Lastly, although a benchmark of .80 is usually considered to reflect a desirable power level in *a priori* power analyses, observing power lower than .80 in *post hoc* power analyses does not mean that important effects have necessarily been missed. It only means that researchers' chances of missing these effects are higher than

20%. As such, unless one observes effects associated with a moderate-to-large effect size that are non-statistically significant and associated with large standard errors, power should not be an issue. After years of highlighting the need to stop reifying statistical significance, it would be a shame to start reifying power analysis using unnecessarily rigid golden rules.

References

Chen, F., Bollen, K.A., Paxton, P., Curran, P.J., & Kirby, J. (2001). Improper solutions in structural models: Causes, consequences, strategies. *Sociological Methods & Research, 29,* 468-508.

de Winter, J.C.F., Dodou D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *44*, 147-181.

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 Ed., Vol. 2* (pp. 1044-1073). Wiley.

Myers, N.D., Ahn, S., & Jin, Y. (2011). Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: A Monte Carlo approach. *Research Quarterly for Exercise and Sport*, *82,* 412-423.

Myers, N.D., Celimli, S., Martin, J.J., Hancock, G.R. (2016). Sample size determination and power estimation in structural equation modeling. In N. Ntoumanis & N.D. Myers (Eds.), *An Introduction to Intermediate and Advanced Statistical Analyses for Sport and Exercise Scientists* (pp. 267-284). Chichester, UK: Wiley.

Myers, N.D., Ntoumanis, N., Gunnell, K.E., Gucciardi, D.F., & Lee, S. (2018). A review of some emergent quantitative analyses in sport and exercise psychology. *International Review of Sport and Exercise Psychology, 11*, 70-100.