Morin, Alexandre J. S.; Arens, A. Katrin; Tran, Antoine; Caci, Hervé

Exploring sources of construct-relevant multidimensionality in psychiatric measurement. A tutorial and illustration using the Composite Scale of Morningness

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
International journal of methods in psychiatric research 25 (2016) 4, S. 277-288, 10.1002/mpr.1485

Bitte verwenden Sie in der Quellenangabe folgende URN oder DOI /
Please use the following URN or DOI for reference:
urn:nbn:de:0111-dipfdocs-201733
10.25657/02:20173

https://nbn-resolving.org/urn:nbn:de:0111-dipfdocs-201733
https://doi.org/10.25657/02:20173

Exploring Sources of Construct-Relevant Multidimensionality in Psychiatric Measurement: A Tutorial and Illustration using the Composite Scale of Morningness

Alexandre J.S. Morin, Institute for Positive Psychology and Education, Australian Catholic University, Australia

A. Katrin Arens, German Institute for International Educational Research, Germany

Antoine Tran, Hôpitaux Pédiatriques de Nice CHU Lenval, France

Hervé Caci, Hôpitaux Pédiatriques de Nice CHU Lenval, France

Acknowledgements
This study is recorded on clinicaltrials.gov under NCT01260792 and was funded by a grant from the French Ministry of Health awarded to the last author (*Programme Hospitalier de Recherche Clinique National 2009*). This article was also made possible by a grant from the Australian Research Council (DP140101559) awarded to the first author and was prepared when the second author was a visiting scholar at the Institute for Positive Psychology and Education, Australia. The authors are grateful to Eric Fontas for his help in setting up the study procedures, Vanina Oliveri and Kevin Dollet for their sustained efforts to monitor the study and collect the data, to the *Inspection Académique des Alpes-Maritimes* and the *Rectorat des Alpes-Maritimes et du Var* for their valuable support, and to the teachers, pupils and parents for participating in this study.

This is the prepublication version of the following manuscript:

Morin, A.J.S., Arens, A.K., Tran, A., & Caci, H. (2016). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the Composite Scale of Morningness. *International Journal of Methods in Psychiatric Research, 25*, 277–288. http://doi.org/10.1002/mpr.1485

Abstract

This paper illustrates a psychometric approach of broad relevance to psychiatric research instruments. Many instruments include indicators related to more than one source of true-score variance due to the: (a) assessment of conceptually adjacent constructs; (b) the presence of a global construct underlying answers to items designed to assess multiple dimensions. Exploratory structural equation modelling (ESEM) is naturally suited to the investigation of the first source, whereas bifactor models are particularly suited to the investigation of the second source. When both sources are present, bifactor-ESEM becomes the model of choice. To illustrate this framework, we use the responses of 1,159 adults (655 female, 504 male, M_{age} = 41.84) who completed the French Version of the Composite Scale of Morningness (CSM). We investigate the factor structure of the CSM, test the relations between CSM factors and Body-Mass Index, and verify the measurement invariance of the model across gender and age groups.

Key words: Exploratory structural equation modelling (ESEM), bifactor models, multidimensionality, diurnal preference, chronotype, Composite Scale of Morningness (CSM).

In psychiatric, epidemiological or biomedical research, a key question is whether unobservable constructs such as personality traits (e.g., Neuroticism, Extraversion), Internalizing Disorders (e.g., anxiety disorders such as Obsessive-Compulsive Disorder, mood disorders such as depression) or Externalizing Disorders (e.g., Attention-Deficit Hyperactivity Disorder [ADHD], Conduct Disorder) exist as a unitary construct including specificities, or represent a collection of correlated/comorbid facets without a common core (Morin et al., in press-b). For example, the DSM-V defines ADHD by a core set manifestations leading to the main diagnosis, and individual specificities characterizing subtypes. Thus, a generic (G) core component of ADHD should co-exist with more specific (S) symptoms (Martel et al., 2011). Similar observations have been previously made for a multitude of constructs such as psychosis (Reininghaus et al., 2013), Internalizing Disorders (Simms et al., 2008), Quality of Life (Reise et al., 2007), or Intelligence (Gignac and Watkins, 2013).

Correlated Constructs, or a Global Construct with Specificities

Psychometrically, the question of whether indicators (questionnaire items, measures, etc.) better depict correlated constructs or a global construct with specificities can be verified by contrasting alternative measurement models. Exploratory or Confirmatory Factor Analyses (EFA and CFA) implicitly assume the presence of separate inter-related dimensions. Conversely, higher-order CFA (H-CFA) directly assesses the presence of a global construct. In H-CFA, indicators are used to define “first-order” factors, themselves used to define a “higher-order” factor (Rindskopf and Rose, 1988). However, H-CFA are limited by their reliance on rigid implicit assumptions (Chen et al., 2006; Jennrich and Bentler, 2011; Reise, 2012). More precisely, H-CFA assume that the relation between each indicator and the higher-order factor is reflected by the combination of the loading of this indicator on a first-order factor, and the loading of this first-order factor on the higher-order factor (a constant as far as the indicators associated with a single first-order factor are concerned). Furthermore, first-order factors reflect a combination of the variance explained by the higher-order factor and the specific variance remaining unexplained by the higher-order factor, creating redundancies between the first-order and higher-order factors. In H-CFA, the disturbances of the first-order factors reflect their specificity remaining unexplained by the higher-order factor. The relations between indicators and these disturbances are also indirect and characterized by the combination of the loadings of the indicators on their first-order factor with a constant for all indicators associated with a single first-order factor. H-CFA models thus rely on stringent proportionality constraints, assuming that the ratio of global/specific variance is exactly the same for all indicators associated with a first-order factor (Jennrich and Bentler, 2011; Reise, 2012). Although these constraints introduce parsimony, they are unlikely to hold in most situations (Reise, 2012; Yung et al., 1999).

Bifactor-CFA models (B-CFA) provide an alternative to H-CFA (Chen et al., 2006). In a *f*-factor B-CFA, one Global (G) factor and *f*-1 Specific (S) factors are used to explain the covariance among a set of *n* indicators. The indicators' loadings on the G-factor and on one of *f*-1 S-factors are estimated while the other loadings are constrained to be zero, and all factors are set to be orthogonal (uncorrelated). B-CFA partitions the total covariance into a G component underlying all indicators, and *f*-1 S components reflecting the residual covariance not explained by the G-factor. Bifactor models directly test the presence of a global construct underlying all indicators (G-factor) and whether this global construct co-exists with meaningful specificities (S-factors), and are able to do so without imposing restrictive proportionality constraints (Chen et al., 2006; Reise, 2012). Furthermore, Jenrich and Bentler (2011) showed that H-CFA models were typically unable to recover the structure of data generated according to bifactor specifications, whereas B-CFA properly recovered H-CFA structures.

Multiple Sources of True Score Variance

B-CFA explicitly accommodates psychometric multidimensionality in the indicators by relaxing the independent cluster assumption (ICM) of CFA according to which each indicator is assumed to correspond to a single factor. Psychometric multidimensionality occurs when indicators are associated with more than one construct, or sources of true score variance (Morin et al., in press-a). Psychometric indicators, be they self-reported, informant-reported, or emerging from structured diagnostic interviews, are very seldom perfectly pure construct indicators. This recognition of the inherently imperfect nature of indicators forms the basis of classical test theory (CTT; Nunnally and Bernstein, 1994), although all implications of this recognition have not been equally well integrated in research. In CTT, ratings are assumed to reflect a combination of true score variance and random measurement error (estimated in reliability analyses). By definition, "random" measurement error is unrelated to other constructs, leading to its absorption within the indicators' uniquenesses in CFA.

CTT further distinguishes among construct-relevant and construct-irrelevant forms of true score variance, a distinction covered in discussions of validity. This distinction makes it obvious that indicators are expected to include at least some degree of association with other constructs. When looking at this issue from the perspective of a single construct, the portion of true score variance that is unrelated to the target construct is simply interpreted as reflecting the imperfect validity of the ratings. However, because this portion still reflects true score variance, it also reflects a form of validity in the assessment of the other constructs to which it is associated – something that only becomes obvious when multiple constructs are simultaneously assessed. For example, using complicated words like "bitterness" or "fallacious" in a measure for children is likely to induce random measurement error due to the need to guess the meaning of the word, producing higher uniquenesses (lower reliability). However, even when completely reliable, ratings of insomnia are likely to present significant levels of true score (i.e., valid) associations with multiple constructs such as depression, anxiety, or drug abuse.

Above, we discussed one process through which indicators might be validly associated with more than one form of true score variance (Morin et al., in press-a) due to the simultaneous assessment of a more global construct (e.g., Intelligence; ADHD) coexisting with specificities (e.g., vocabulary; hyperactivity). Bifactor models are required to directly investigate this possibility (Chen et al., 2006; Reise, 2012). Indeed, if data simulated according to a bifactor-CFA was analysed using ICM-CFA, the unmodelled G-factor would be absorbed through an inflation of the factor correlations, calling into question the discriminant validity of the factors (Morin et al., in press-a).

It is also typical for indicators to present construct-relevant associations with more than one source of true score variance located at the same conceptual level, particularly in instruments designed to assess conceptually-related and partially overlapping domains, such as inattention and hyperactivity (ADHD), or depression and anxiety (Internalizing Disorders). This second form of construct-relevant multidimensionality is typically expressed through cross-loadings in EFA but is constrained to be zero in ICM-CFA, H-CFA, or B-CFA. The simple observation that many indicators are inherently expected to present meaningful associations to multiple sources of true score variance shows that ICM requirement for pure indicators relies on an inherently flawed logic.

In sum most psychometric indicators are likely to present at least some level of systematic association with other constructs. Although "pure" indicators may exist, we surmise that such indicators remain at best a convenient fiction (Marsh et al., 2014; Morin et al., in press-a). Simulation studies have clearly demonstrated that, even when small (i.e., as low as .100) and substantively meaningless cross-loadings are present in the population model but ignored in ICM-CFA models, the factor correlations will tend to be substantially biased (Asparouhov and Muthén, 2009; Marsh et al., 2013; Morin et al., in press-a; Schmitt and Sass, 2011). Although B-CFA models relax ICM assumptions to some extent, they still ignore cross-loadings, which tends to result in inflated estimates of the variance attributed to the global factor (Morin et al., in press-a; Murray and Johnson, 2013). These studies also show that when the population model meets ICM assumptions, relying on models allowing for the estimation of cross-loadings (e.g., EFA) will nevertheless result in unbiased estimates of factor correlations. Going back to the flawed argument that cross-loadings "change" the nature of the constructs, these results rather show that it is the exclusion of cross-loadings that modifies the meaning of the constructs.

Reviving Exploratory Factor Analyses (EFA)

The foregoing arguments seem to support the revival of classical EFA. Unfortunately, EFA has been superseded by the methodological advances associated with CFA/SEM (e.g., goodness-of-fit, invariance, predictions, etc.) and the erroneous assumption that EFA was not confirmatory. However, the only "*critical difference between EFA and CFA is that all cross loadings are freely estimated in EFA. Due to this free estimation of all cross loadings, EFA is clearly more naturally suited to exploration than CFA. However, statistically, nothing precludes the use of EFA for confirmatory purposes*" (Morin et al., 2013, p.396). However, because classical EFA models rely on the free estimation of all loadings and cross loadings, they have also been criticized for the fact that this free estimation of multiple parameters may yield overfitting the data and create an undue level of sensitivity to random variations across different data sets. However, EFA has recently been integrated with CFA/SEM into the Exploratory Structural Equation Modelling (ESEM; Asparouhov and Muthén, 2009) framework, making most methodological advances typically reserved to CFA/SEM available for EFA (Marsh et al., 2013, 2014; Morin et al., 2013). In particular, the use of goodness-of-fit indices adjusted for parsimony makes it easier to compare more parsimonious CFA with EFA models while taking into account the fact that EFA models rely on the estimation of many additional parameters. The development of target rotation also makes it possible to use a fully confirmatory approach to EFA (Asparouhov and Muthén, 2009; Browne, 2001) through which cross-loadings are freely estimated but "targeted" *a priori* to be as close to zero as possible. Finally, bifactor rotations (Jennrich and Bentler, 2011), including bifactor target rotation (Reise, 2012; Reise et al., 2011), have recently been developed for the estimation of bifactor-ESEM (B-ESEM) models. Finally, ESEM makes it possible to directly assess the extent to which an EFA solution can be generalizable across samples, providing a more systematic way to directly test the sensitivity of the solution to random sample variations.

Taken together, these developments form an overarching framework for the investigation of two sources of construct-relevant psychometric multidimensionality likely to be present in many psychiatric measures. The assessment of hierarchically-organized construct calls for bifactor models, whereas the assessment of conceptually-adjacent constructs calls for ESEM. However, bifactor models are likely to express unmodelled cross-loadings through an inflated G-factor, whereas ESEM models are likely to express an unmodelled G-factor through inflated cross-loadings. B-ESEM models are thus most suitable when a measure includes hierarchically-organized and conceptually-adjacent constructs.

In this study, we illustrate this B-ESEM framework using self-reports on the Composite Scale of Morningness (CSM) (Caci et al., 2005, 2009), a short (13-item) measure of Chronotype or diurnal preference (an inter-individual difference related to the time of day where a person is the most alert and awake, and to preferences for early or late awakening). Picking a short (13 items) and simple scale helps to keep the illustration (for which we provide annotated Mplus input codes in the Online Supplements) as simple as possible, while demonstrating the broad relevance of this framework for psychiatric measurement. We provide theoretical background on the CSM in the online supplements.

METHOD

Participants and Material

This illustration uses data obtained from the parents of the youth involved in the ChiP-ARD study conducted in 2010-2011 in 20 kindergarten schools, 30 primary schools, 14 secondary schools from Southern France (Caci et al., 2014, in press). Schools were randomly drawn from all public schools located in the greater Nice area, and invited to participate until a number of schools sufficient to reach a sample size of approximately 1,000 participants, equally distributed by age group, had been recruited. Teachers were then individually invited to participate, and those who agreed sent consent forms to the parents of a randomly selected subset of students from their classes (2 to 4 students for each class). Through this procedure, 941 students were finally included in the study. In the present study, we use data from the parental questionnaires, including the French CSM (Caci et al., 2005, 2009) and self-reported height and weight. Taking into account the prevalence of single parent families, reconstituted families, and families where parents do not fluently speak French based on the 2009 Census for France (http://www.insee.fr/en/default.asp), our expected sample was 1411 (1.5. parent per family). In total, 1166 parents (82.63%) returned completed CSM questionnaires. Seven pregnant women were excluded due to the impact of pregnancy on sleep cycles and BMI. This sample includes 1,159 parents (22 to 65 years old; M_{age} = 41.84; M_{BMI} = 23.53), including 655 women (56.51%; M_{age} = 40.84; M_{BMI} = 22.27) and 504 males (43.49%; M_{age} = 43.12; M_{BMI} = 25.15). Compared to 2009 Census data for the city of Nice, this sample tended to be slightly more educated, but remained quite representative of the general adult population of Nice (for additional details, see Caci et al., 2014). This study is supported by the Commissioner of Education and the Department of Education, complied with ethical prescriptions for French medical research, and data management procedures were approved by the Commission Nationale Informatique et Liberté.

Statistical analyses

Measurement models were estimated using Mplus 7.2 (Muthén and Muthén, 2012) robust weight least square estimator (WLSMV) which outperforms Maximum Likelihood for ordered-categorical indicators with 5 or less answer categories (Beauducel and Herzberg, 2006; Finney and DiStefano, 2006). CSM items (see Online Supplements) were recoded prior to the analyses so that a higher score reflected morning preference. We successively estimated ICM-CFA, B-CFA, ESEM, and B-ESEM models based on the revised CSM 3-factor structure (see Online Supplements). Models based on the original factor structure were also estimated, but the results fully supported the superiority of the revised factor structure. ESEM was estimated using target rotation, while B-ESEM was estimated using bifactor-target rotation (Reise, 2012; Reise et al., 2011). ICM-CFA and B-CFA constrained all cross-loadings to be exactly zero, while ESEM and B-ESEM targeted all cross-loadings to be as close to zero as possible. In both B-CFA and B-ESEM, all indicators were allowed to load on a global G-factor and on a specific a priori S-factor. BMI was then integrated to these models as an outcome predicted by the estimated factors.

Composite reliability was calculated using McDonald's $\omega = (\Sigma|\lambda_i|)^2 / ([\Sigma|\lambda_i|]^2 + \Sigma\delta_{ii})$ where λ_i are the factor loadings and δ_{ii}, the uniquenesses (McDonald, 1970). Compared with alpha, ω has the advantages of being model-based and of taking into account the strength of association between indicators and constructs (λ_i) as well as item-specific measurement errors (δ_{ii}) (Sijtsma, 2009).

The final model was submitted to tests of measurement invariance across gender (male versus females), age groups (adults younger than 40 years versus older than 40 years), and combinations (younger males, older males, younger females, older females). These tests followed the typical sequential invariance strategy (Meredith, 1993) adapted for ordered-categorical indicators (Guay et al., in press; Morin et al., 2011): (i) configural invariance, (ii) metric/weak invariance (invariance of the factor loadings); (iii) scalar/strong invariance (loadings and thresholds); (iv) strict invariance (loadings, thresholds and uniquenesses), (v) invariance of the latent variances-covariances (loadings, thresholds, uniquenesses and variances-covariances), and (vi) latent means invariance (loadings, thresholds, uniquenesses, variances-covariances and latent means).

The fit of all models was evaluated using the WLSMV χ^2, the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA) and its 90% confidence interval (Hu and Bentler, 1999; Yu, 2002). Values greater than .900 and .950 for CFI and TLI, and lower than .080 and .060 for the RMSEA are respectively indicative of adequate and excellent model fit. Fit improvement was evaluated using the MPlus DIFFTEST function (MD$\Delta\chi^2$; (Asparouhov and Muthén, 2006; Muthén, 2004). Because χ^2 and MD$\Delta\chi^2$ tend to be oversensitive to sample size and to minor misspecifications, additional indices were used in tests of invariance (Chen, 2007; Cheung and Rensvold, 2002): a CFI diminution of .010 or less and a RMSEA augmentation of .015 or less between a model and the preceding model indicate that the measurement invariance hypothesis should not be rejected. With WLSMV, χ^2 values are not exact, but adjusted to obtain a correct *p* value. This explains why χ^2 and CFI can be nonmonotonic with model complexity. CFI improvements should thus be interpreted as random. In contrast, both the TLI and the RMSEA are adjusted for the parsimony of the model and, as such, can increase with invariance constraints.

RESULTS

The goodness-of-fit indices of the alternative models are reported in Table 1. Results show that ICM-CFA provides an unacceptable level of fit to the data (CFI and TLI < .900; RMSEA > .100). B-CFA and ESEM yield a clearly improved level of fit although they both remain marginal according to some indices (TLI < .950; RMSEA > .080). These results suggest that both sources of construct-relevant psychometric multidimensionality may be present in CSM ratings. Indeed, B-ESEM clearly provides the best fit to the data. Thus, based on purely statistical criteria, B-ESEM should be retained. However, no analysis should be conducted in disconnection from theory, expectations, and a detailed examination of parameter estimates (Marsh et al., 2004; Morin et al., in press-a).

Table 2 presents the parameter estimates from all models. In ICM-CFA, all factors appear well defined by high and significant factors loadings (.50-.96; M =.73) and satisfactory composite reliability (ω =.77-.88). However, the fact that this model results in such a poor level of fit to the data suggests that it fails to properly represent the underlying structure of the data. Furthermore, ICM-CFA factor correlations (.46-.73; M =.63) appear high enough to call into question the discriminant validity of some factors, suggesting that CSM ratings may include unmodelled multidimensionality. ESEM reveals a substantial reduction of the factor correlations (.30-.57; M =.45) while all factors remain clearly defined (.50-.96; M = .73) and reliable (ω = .80-.87). However, although most cross-loadings remain small (|.02-.36|; M = .14), some are high enough (> .30 for Items 3 and 9) to suggest that another source of unmodelled multidimensionality may be present, explaining the marginal fit of this model (TLI <.950; RMSEA >.080).

This hypothesis is readily confirmed when comparing B-CFA with ICM-CFA. Apart from providing a better fit to the data, B-CFA also results is a well-defined G-factor (λ =.16-.77; M =.59; ω =.92). Indeed, apart from two items associated with the *Activity Planning* S-factor that present a lower loading on the G-factor (item 2: λ =.34, "…*at what time would you go to bed if you were entirely free to plan your evening?*"; item 7: λ =.16, "*At what time in the evening do you feel tired …?*"), the remaining items all have fully satisfactory loadings on the G-factor (.50 to .77; M = .65) reflecting global diurnal preference. Interestingly, these two items present high loadings on their corresponding S-factor "Activity Planning" (.61 and .85), whereas the two remaining indicators of this S-factors present much lower loadings (Item 9: λ =.36, "*One hears about morning and evening types of people …?*"; Item 13: λ = .36, "*… a morning or evening active individual?*"). This suggests that these two items (9 and 13) mainly reflect global diurnal preference and only present a low level of specificity once their association with the G-factor is taken into account. In contrast, the third S-factor mainly appears to represent the specificity associated with items 2 and 7, reflecting *Bedtime Preference* (rather than *Activity Planning*). This S-factor shares similarity with the *Evening Activities* factor from the original CSM model, while still presenting a significant, albeit small, association with items 9 and 13. Interestingly, this S-factor still presents satisfactory reliability (ω =.75).

The results further reveal that the first S-factor (*Morning Affect*) retains a meaningful level of specificity (λ =.37-.72; M =.59; ω =.78), while the second S-factor (*Time of Rising*) apparently retains no meaningful specificity (|.07-.52|; M =.23; ω =.43), and mainly serve control for the limited level of residual covariance present in these items once the G-factor is taken into account. Although this B-CFA model is interesting, the fit of this model is marginal (TLI <.950; RMSEA >.080) and below the fit of ESEM (ΔCFI, ΔTLI, ΔRMSEA all $\geq$.100), suggesting unmodelled cross-loadings.

The B-ESEM solution supports this assertion. First, the fit of this model is fully satisfactory and clearly superior to the fit of all alternative models. Second, the pattern of target loadings mimics the B-CFA results, showing: (a) a G-factor that is well defined by most items (λ =.50-.84; M =.62; ω =.92) apart from items 2 (.29) and 7 (.24) which together define a *Bedtime Preference* S-factor (λ =.72 and .73; versus .26 and .25 for items 9 and 13; ω =.74); (b) a well-defined *Morning Affect* S-factor (.41-0.69; M =.58; ω =.80); (c) a weakly defined *Time of Rising* S-factor (.04-.47; M =.31; ω =.52). However, although the reliability of the *Time of Rising* S-factor remains suboptimal (which is not an issue in latent models controlling for reliability), the level of specificity associated with this S-factor is higher than in B-CFA. Finally, the cross-loadings remain smaller (|.01-.23|; M =.09) than in ESEM (|.02-.36|; M =.14), suggesting that construct-relevant multidimensionality initially absorbed in the cross-loadings now serves to map the G-factor.

Associations with BMI

To illustrate the impact of suboptimal measurement models for predictive analyses, we present the results of analyses in which CSM factors from the four alternative models are used to predict BMI (Table 3). The ICM-CFA results are highly similar to the ESEM results, and the B-CFA results are highly similar to the B-ESEM results. This similarity is likely due to the fact that the factors remain equally well-defined in the ICM-CFA/ESEM, and B-CFA/B-ESEM, and the fact that cross-loadings are small. Three of the four models (ICM-CFA, ESEM, and B-ESEM) result in comparable estimates of the percentage of explained variance in BMI levels (approximately 3%). However, the results show that retaining a suboptimal model results either in highly different substantive conclusions (ICM-CFA, ESEM) or a substantially reduced percentage of explained variance (closer to 2% for B-CFA). Thus, when the global diurnal preference factor is "absorbed" through first order factor correlations (ICM-CFA) or cross-loadings (ESEM), the results suggest that all three CSM factors significantly predict BMI. However, when global chronotype is explicitly taken into account (B-ESEM, B-CFA), neither global diurnal preference, nor the *Bedtime Preference* S-factor share any relations with BMI. Rather, more positive *Morning Affect* predicts slightly lower BMI levels, whereas an earlier *Time of Rising* predicts higher BMI levels.

Measurement Invariance

The results from the tests of the measurement invariance conducted on the best-fitting B-ESEM solution are presented in Table 1. Although all χ^2 and some $\Delta\chi^2$ are significant, the goodness-of-fit indices indicate fully satisfactory model fit at each stage. Furthermore, changes in goodness-of-fit indices never decrease more than the recommended guidelines when equality constraints are imposed on the loadings, intercepts, uniquenesses, and factor variances-covariances. The TLI and RMSEA even revealed an improvement in fit at some steps. Strict measurement invariance of the CSM is thus supported across gender, age groups, and combinations, as well as the invariance of the latent variance-covariance matrix. The results also suggest the presence of latent mean differences, particularly in the combined model (Table 4). Although the four groups do not differ on the *Diurnal Preference* G-factor, older females tend to present a more positive *Morning Affect* than younger females, and males tend to prefer an earlier *Time of Rising* than females. Although this S-Factor presents a lower level of specificity, the fact that these comparisons are based on *latent* means indicates that they are perfectly reliable. Finally, the results show that men have a later *Bedtime Preference* than females, but that older males prefer getting into bed earlier than younger males.

DISCUSSION

In psychiatric, epidemiological and biomedical research, the factor validity of psychiatric instruments is typically assessed using first-order or higher-order CFA or EFA. We argued that bifactor models provide a more flexible, realistic, and meaningful representation of the data whenever these dimensions are assumed to reflect a global underlying construct. We also discussed how the assessment of conceptually-adjacent dimensions may lead to psychometric complexity due to the unrealism of the expectation that indicators should provide a perfect reflection of a single construct. Rather, many indicators correspond to more than one source of true score variance, leading them to present significant associations with more than one construct. We argued that the first source of construct-relevant psychometric multidimensionality naturally calls for bifactor models (Reise, 2012), whereas the second source rather calls for ESEM (Marsh et al., 2014). Finally, bifactor-ESEM appears to be preferable when both sources of construct-relevant psychometric multidimensionality are present (Morin et al., in press-a). More importantly, the failure to properly consider these sources of construct-relevant multidimensionality might induce potentially severe biases in terms of both assessment and prediction (Marsh et al., 2013, 2014; Murray and Johnson, 2013; Schmitt and Sass, 2011).

This manuscript presented this overarching bifactor-ESEM framework of broad relevance to psychiatric, epidemiological and biomedical research. The implementation of this framework was illustrated while using a WLSMV estimation process allowing for a proper representation of the ordered-categorical nature of response scales frequently used in psychiatric diagnostic ratings.

The application of this framework starts with a comparison of ICM-CFA and ESEM to test for the presence of multidimensionality due to conceptually-adjacent constructs. Because bifactor models tend to absorb unmodelled cross-loadings through inflated global factors, it is critical to start with a comparison of ICM-CFA and ESEM. In this comparison, observing substantially reduced factor correlations, better fit indices, substantive meaningfulness, and small or easy to explain cross-loadings argues in favour of ESEM (Marsh et al., 2013, 2014; Morin et al., 2013, in press-a). In particular, the observation of multiple cross-loadings of a reasonable magnitude ($\geq .10$ or even $\geq .20$) in the ESEM solution is particularly important and suggests that a global construct might be present in the data.

As long as there are reasons to suspect that a global construct might be present, the second step is to test this possibility by comparing ICM-CFA and B-CFA. Over and above the observation of better-fit indices associated with B-CFA, a critical element is the presence of a well-defined G-factor. Whenever this is the case, a bifactor representation of the data appears justified. Although it is not critical for all S-factors to be equally well-defined – S-factors may sometimes be included to control for residual specificities shared among subsets of indicators over and above their association with the G-factor – a true bifactor representation should typically result in at least some well-defined S-factors. Otherwise, a single-factor model should be seriously considered. Undefined S-factors should simply not be interpreted as having a substantive meaning.

When both sources of construct-relevant psychometric multidimensionality appear to be present based on substantive expectations and the results from the previous steps, a B-ESEM representation should be pursued. The adequacy of this representation would be supported by the observation of: (a) improved goodness-of-fit indices; (b) a well-defined global factor; (c) relatively small cross-loadings, ideally smaller than those associated with the ESEM model. Although our results supported a B-ESEM representation, we do not claim that this framework should be blindly applied to all measures, or that B-ESEM will always prove superior. As in any statistical analyses, there is a need to combine substantive theory, expectations, common-sense, and proper statistics in order to achieve an adequate representation of the data (Morin et al., in press-a). However, we expect that the bifactor-ESEM combination may prove to be relevant for a substantial number of applications using complex multidimensional measures. It is thus our recommendation that the sequential process described here (i.e., contrasting ICM-CFA versus ESEM, ICM-CFA versus B-CFA, and then all of these models versus B-ESEM) should be routinely applied to studies of complex instruments.

The framework described here relies on variable-centred analyses, providing results reflecting a synthesis of the relations observed in the total sample. In contrast, person-centred methods aim to identify subgroups of participants (i.e., profiles), which qualitatively and quantitatively differ from one another on a configuration of indicators (Morin and Marsh, 2015). Hybrid approaches provide a way to represent similar forms of construct-relevant multidimensionality through the estimation of a variable-centred factor (reflecting a global tendency shared among indicators) and person-centred profiles (reflecting specific areas of strength and weaknesses over and above this global tendency) from the same set of indicators (Morin and Marsh, 2015). Importantly, this hybrid framework can be used to conduct even more refined explorations of the underlying structure (categorical, continuous, ordinal, etc.) and dimensionality of psychiatric constructs (for details, see Clark et al., 2013; Masyn et al., 2010). However, this approach requires the estimation of complex and computer-intensive models with a known tendency to converge on improper solutions or not to converge at all. For this reason, most applications of this hybrid framework uses scale scores (i.e., the sum/average of items used to assess a specific dimension), or factor scores from preliminary measurement models (e.g., Morin and Marsh, 2015) as indicators. Estimated in this manner, hybrid models thus assume that these scale or factor scores provide a proper synthesis of the underlying structure of participants' responses. In this context, the bifactor-ESEM framework presented here appears to represent a critical first step in the application of these potentially richer hybrid methodologies. Future statistical research would do well to examine more attentively the possible impact of misspecifying the factor structure of an instrument when scale/factor scores from this instrument are used in person-centred applications.

REFERENCES

Asparouhov T, Muthén B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling* **16**(3): 397-443, DOI: 10.1080/10705510903008204.

Asparouhov T, Muthén BO. (2006). Robust chi-square difference testing with mean and variance adjusted test statistics. Los Angeles, CA: Muthén and Muthén; www.statmodel.com/examples/webnote.shtml - web10.

Beauducel A, Herzberg PY. (2006). On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling* **13**(2): 186-203, DOI: 10.1207/s15328007sem1302_2.

Browne MW. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research* **36**(1): 111-150, DOI: 10.1207/S15327906MBR3601_05.

Caci H, Adan A, Bohlé P, Natale V, Pornpitakpan C, Tilley A. (2005). Transcultural properties of the Composite Scale of Morningness: the relevance of the 'Morning Affect' factor. *Chronobiology*

International **22**(3): 523-540, DOI: 10.1081/CBI-200062401.

Caci H, Deschaux O, Adan A, Natale V. (2009). Comparing three morningness scales: age and gender effects, structure and cut-off criteria. *Sleep Medicine* **10**(2): 240-245, DOI: 10.1016/j.sleep.2008.01.007.

Caci H, Morin AJS, Tran A. (2014). Prevalence and correlates of Attention Deficit/Hyperactivity Disorder in adults from a French community sample. *Journal of Nervous and Mental Disease* **202**(4): 324-332, DOI: 10.1097/NMD.0000000000000126.

Caci H, Morin AJS, Tran A. (in press). Teacher ratings of the ADHD-RS IV in a community sample: Results from the ChiP-ARD study. *Journal of Attention Disorders*, DOI: 10.1177/1087054712473834.

Chen FF. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling* **14**(3): 464-504, DOI: 10.1080/10705510701301834.

Chen FF, West SG, Sousa KH. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research* **41**(2): 189-225, DOI: 10.1207/s15327906mbr4102_5.

Cheung GW, Rensvold RB. (2002). Evaluating goodness-of fit indexes for testing measurement invariance. *Structural Equation Modeling* **9**(2): 233-255, DOI: 10.1207/S15328007SEM0902_05.

Clark SL, Muthén BO, Kaprio J, D'Onofrio BM, Viken R, Rose RJ. (2013). Models and strategies for factor mixture analysis: An example concerning the structure of underlying psychological disorders. *Structural Equation Modeling* **20**(4): 681-703, DOI: 10.1080/10705511.2013.824786.

Finney SJ, DiStefano C. (2006). Non-normal and categorical data in structural equation modeling. In GR Hancock, RO Mueller (eds.) *Structural Equation Modeling: A Second Course*. Greenwich,CT: IAP. pp 269-314.

Gignac GE, Watkins MW. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research* **48(5)**: 639-662, DOI: 10.1080/00273171.2013.804398.

Guay F, Morin AJS, Litalien D, Valois P, Vallerand RJ. (in press). An application of exploratory structural equation modeling to evaluate the construct validity and the gender invariance of the

Academic Motivation Scale. *The Journal of Experimental Education*, DOI: 10.1080/00220973.2013.876231.

Hu LT, Bentler PM. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* **6**(1): 1-55, DOI: 10.1080/10705519909540118.

Jennrich RI, Bentler PM. (2011). Exploratory Bi-factor Analysis. *Psychometrika* **76**(4): 537-549, DOI: 10.1007/s11336-011-9218.4.

Marsh HW, Hau K-T, Wen Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999). *Structural Equation Modeling* **11**(3): 320-341, DOI: 10.1207/s15328007sem1103_2.

Marsh HW, Lüdtke O, Nagengast B, Morin AJS, Von Davier M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—Camouflaging misspecification with item parcels in CFA models. *Psychological Methods* **18**(3): 257-284, DOI: 10.1037/a0032773.

Marsh HW, Morin AJS, Parker P, 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, DOI: 10.1146/annurev-clinpsy-032813-153700.

Martel MM, Roberts B, Gremillion M, von Eye A, Nigg JT. (2011). External validation of the bifactor model of ADHD: explaining heterogeneity in psychiatric comorbidity, cognitive control, and personality trait profiles within DSM-IV ADHD. *Journal of Abnormal Child Psychology* **39**(8): 1111-1123, DOI: 10.1007/s10802-011-9538-y.

Masyn K, Henderson C, Greenbaum P. (2010). Exploring the latent structures of psychological constructs in social development using the Dimensional-Categorical Spectrum. *Social Development* **19**(3): 470–493, DOI: 10.1111/j.1467-9507.2009.00573.x.

McDonald RP. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *Bristish Journal of Mathematical and Statistical Psychology* **23**(1): 1-21, DOI: 10.1111/j.2044-8317.1970.tb00432.x.

Meredith W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*

58(4): 525-543, DOI: 10.1007/BF02294825.

Morin AJS, Arens AK, Marsh HW. (in press-a). A Bifactor Exploratory Structural Equation Modeling Framework for the Identification of Distinct Sources of Construct-Relevant Psychometric Multidimensionality. *Structural Equation Modeling*.

Morin AJS, 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 & Exercise* **12**(5): 540-554, DOI: 10.1016/j.psychsport.2011.04.003.

Morin AJS, Marsh HW. (2015). Disentangling shape from levels effects in person-centred analyses: An illustration based university teacher multidimensional profiles of effectiveness. *Structural Equation Modeling* **22**(1): 39-59, DOI: 10.1080/10705511.2014.919825.

Morin AJS, Marsh HW, Nagengast B. (2013). Exploratory Structural Equation Modeling. In GR Hancock, RO Mueller (eds.) *Structural Equation Modeling: A Second Course*. Greewich,Connecticut: IAP. pp 395-436.

Morin AJS, Moullec G, Maïano C, Layet L, Just J-L, Ninot G. (2011). Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in French Clinical and Non-Clinical Adults. *Epidemiology and Public Health [Revue d'Épidémiologie et de Santé Publique]* **59**(5): 327-340, DOI: 10.1016/j.respe.2011.03.061.

Morin AJS, Tran A, Caci H. (in press-b). Factorial Validity of the ADHD Adult Symptom Rating Scale in a French community sample: Results from the ChiP-ARD study. *Journal of Attention Disorders*, DOI: 10.1177/1087054713488825.

Murray AL, Johnson W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence* **41**(5): 407-422, DOI: 10.1016/j.intell.2013.06.004.

Muthén BO. (2004). Mplus Technical Appendices. Los Angeles, CA: Muthén and Muthén; www.statmodel.com/techappen.shtml.

Muthén LK, Muthén BO. (2012). *Mplus User's Guide. Seventh Edition.* Los Angeles, CA: Muthén & Muthén.

Nunnally JC, Bernstein IH. (1994). *Psychometric theory*: McGraw-Hill, Inc.

Reininghaus U, Priebe S, Bentall RP (2013). Testing the psychopathology of psychosis: evidence for a general psychosis dimension. *Schizophrenia Bulletin* **39**(4): 884-95, DOI: 10.1093/schbul/sbr182.

Reise SP. (2012). Invited paper: The rediscovery of bifactor measurement models. *Multivariate Behavioral Research* **47**(5): 667-696, DOI: 10.1080/00273171.2012.715555.

Reise SP, Moore TM, Maydeu-Olivares A. (2011). Target rotations and assessing the impact of model violations on the parameters of unidimensional item response theory models *Educational and Psychological Measurement* **71**(4): 684-711, DOI: 10.1177/0013164410378690.

Reise SP, Morizot J, Hays RD. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research* **16**(1): 19-31, DOI: 10.1007/s11136-007-9183-7.

Rindskopf D, Rose T. (1988). Some theory and application of confirmatory second-order factor analyses. *Multivariate Behavioral Research* **23**(1): 51-67, DOI: 10.1207/s15327906mbr2301_3.

Schmitt TA, Sass DA. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: implications for factor pattern loadings and interfactor correlations. *Educational & Psychological Measurement* **71**(1): 95-113, DOI: 10.1177/0013164410387348.

Sijtsma K. (2009). On the use, misuse, and the very limited usefulness of Cronbach's alpha [Introduction to a special issue]. *Psychometrika* **74**(1): 107-120, DOI: 10.1007/s11336-008-9101-0.

Simms LJ, Grös DF, Watson D, O'Hara M. (2008). Parsing general and specific components of depression and anxiety with bifactor modeling. *Depression and Anxiety* **25**(7): E34–E46, DOI: 10.1002/da.20432.

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

Yung Y-F, Thissen D, McLeod LD. (1999). On the relationship between the higher order factor model and the hierarchical factor model. *Psychometrika* **64**(2): 113-128, DOI: 10.1007/BF02294531.

Table 1. Goodness-of-Fit Statistics of the Alternative Measurement Models

	WLSMV χ^2 (df)	CFI	TLI	RMSEA	90%CI	$\Delta\chi^2$ (df)	ΔCFI	ΔTLI	ΔRMSEA
ICM-CFA	1252.33 (62)*	0.892	0.864	0.129	[0.123 ; 0.135]	—	—	—	—
B-CFA	556.26 (52)*	0.954	0.932	0.091	[0.085 ; 0.098]	—	—	—	—
ESEM	358.66 (42)*	0.971	0.947	0.081	[0.073 ; 0.088]	—	—	—	—
B-ESEM	105.24 (32)*	0.993	0.984	0.044	[0.035 ; 0.054]	—	—	—	—
Sex invariance									
Configural	107.23* (64)	0.996	0.991	0.034	[0.022 ; 0.045]	—	—	—	—
Weak (loadings)	222.50* (100)	0.989	0.983	0.046	[0.038 ; 0.054]	106.05* (36)	-0.007	-0.008	+0.012
Strong (loadings, intercepts)	306.93* (125)	0.984	0.980	0.050	[0.043 ; 0.057]	91.95* (25)	-0.005	-0.003	+0.004
Strict (loadings, intercepts, uniqu.)	336.08* (138)	0.983	0.980	0.050	[0.043 ; 0.057]	41.12* (13)	-0.001	0.000	0.000
Latent Variance-Covariance	242.32* (148)	0.992	0.991	0.033	[0.025 ; 0.041]	11.51 (10)	+0.009	+0.011	-0.017
Latent Means	358.27* (152)	0.982	0.981	0.048	[0.042 ; 0.055]	56.76* (4)	-0.010	-0.010	+0.015
Age invariance									
Configural	110.39* (64)	0.996	0.990	0.036	[0.024 ; 0.047]	—	—	—	—
Weak (loadings)	114.67* (100)	0.999	0.998	0.016	[0.000 ; 0.028]	27.78 (36)	+0.003	+0.008	-0.020
Strong (loadings, intercepts)	138.08* (125)	0.999	0.999	0.014	[0.000 ; 0.025]	25.34 (25)	0.000	+0.001	-0.002
Strict (loadings, intercepts, uniqu.)	177.39* (138)	0.996	0.996	0.022	[0.011 ; 0.031]	33.14* (13)	-0.003	-0.003	+0.008
Latent Variance-Covariance	149.85* (148)	1.000	1.000	0.005	[0.000 ; 0.020]	5.55 (10)	+0.004	+0.004	-0.017
Latent Means	193.34* (152)	0.996	0.996	0.022	[0.011 ; 0.031]	21.49* (4)	-0.004	-0.004	+0.017
Sex × Age invariance									
Configural	167.27* (128)	0.997	0.992	0.033	[0.016 ; 0.046]	—	—	—	—
Weak (loadings)	346.67* (236)	0.990	0.987	0.040	[0.031 ; 0.049]	180.64* (108)	-0.007	-0.005	+0.007
Strong (loadings, intercepts)	464.18* (311)	0.986	0.986	0.041	[0.033 ; 0.049]	131.32* (75)	-0.004	-0.001	+0.001
Strict (loadings, intercepts, uniqu.)	528.32* (350)	0.984	0.986	0.042	[0.035 ; 0.049]	74.58*(39)	-0.002	0.000	+0.001
Latent Variance-Covariance	466.13* (380)	0.992	0.994	0.028	[0.018 ; 0.036]	29.89 (30)	+0.008	+0.008	-0.014
Latent Means	610.52* (392)	0.981	0.985	0.044	[0.037 ; 0.051]	79.37* (12)	-0.011	-0.009	+0.016

Note. ICM = Independent cluster model; CFA = Confirmatory factor analysis; B = Bifactor model; ESEM = Exploratory structural equation modelling; WLSMV: Robust weighted least square estimator using a diagonal weight matrix, and mean- and variance- adjusted test statistics; χ^2 = WLSMV chi square; df = Degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; Δ since previous model; $\Delta\chi^2$: chi square difference test based on the Mplus DIFFTEST function for WLSMV estimation; ESEM were estimated with target oblique rotation; bifactor-ESEM were estimated with bifactor orthogonal target rotation; * $p < 0.01$.

Table 2. Standardized Parameter Estimates from the Alternative Measurement Models.

	ICM-CFA		Bifactor-CFA			ESEM				Bifactor-ESEM				
Items	Factor (λ)	δ	G-Factor (λ)	S-Factor (λ)	δ	Factor 1 (λ)	Factor 2 (λ)	Factor 3 (λ)	δ	G-Factor(λ)	S-Factor 1(λ)	S-Factor 2(λ)	S-Factor 3(λ)	δ
Morning Affect														
Item 3[a]	**0.79**	0.37	**0.57**	**0.37**	0.45	**0.51**	0.34	*-0.03*	0.44	**0.59**	**0.41**	0.21	-0.06	0.44
Item 4	**0.87**	0.25	**0.64**	**0.72**	0.15	**0.96**	-0.11	0.06	0.16	**0.56**	**0.69**	*-0.03*	*0.01*	0.21
Item 5	**0.82**	0.33	**0.56**	**0.60**	0.33	**0.80**	0.11	-0.11	0.30	**0.53**	**0.66**	0.08	-0.12	0.27
Item 12	**0.72**	0.48	**0.50**	**0.53**	0.47	**0.73**	*-0.04*	0.07	0.47	**0.50**	**0.54**	*-0.05*	*0.02*	0.45
Time of Rising														
Item 1	**0.76**	0.42	**0.72**	**0.27**	0.42	*-0.03*	**0.66**	0.22	0.40	**0.61**	0.06	**0.46**	0.20	0.37
Item 6	**0.61**	0.62	**0.64**	***-0.08***	0.58	0.30	**0.40**	*-0.03*	0.63	**0.53**	0.23	**0.19**	-0.08	0.62
Item 8	**0.66**	0.56	**0.68**	***-0.07***	0.53	0.17	**0.36**	0.26	0.40	**0.66**	*0.02*	***0.04***	*0.06*	0.56
Item 10	**0.73**	0.47	**0.67**	**0.52**	0.28	-0.13	**0.80**	0.11	0.38	**0.64**	-0.06	**0.47**	0.05	0.36
Item 11	**0.77**	0.41	**0.74**	**0.21**	0.40	0.15	**0.74**	-0.06	0.36	**0.67**	0.16	**0.41**	-0.09	0.36
Activity Planning														
Item 2	**0.61**	0.63	**0.34**	**0.61**	0.51	*-0.02*	-0.08	**0.77**	0.46	**0.29**	*0.04*	0.17	**0.72**	0.36
Item 7	**0.50**	0.75	**0.16**	**0.85**	0.26	-0.12	-0.20	**0.86**	0.40	**0.24**	-0.13	*0.01*	**0.73**	0.39
Item 9	**0.96**	0.56	**0.77**	**0.36**	0.28	0.14	0.36	**0.53**	0.29	**0.84**	-0.07	*0.01*	**0.26**	0.23
Item 13	**0.74**	0.45	**0.62**	**0.36**	0.49	0.17	0.17	**0.53**	0.49	**0.74**	-0.08	-0.22	**0.25**	0.34
Correlations														
	Factor 2	Factor 3	G-Factor	Factor 2	Factor 3	Factor 2	Factor 3			G-Factor	Factor 2	Factor 3		
Factor 1	0.71	0.46	0.00	0.00	0.00	0.57	0.30			0.00	0.00	0.00		
Factor 2		0.73	0.00		0.00		0.48			0.00		0.00		
Factor 3			0.00							0.00				
Reliability	ω	α	ω	α		ω	α			ω	α			
G-Factor	---		0.92	0.86		---				0.92	0.86			
Factor 1	0.88	0.81	0.78	0.81		0.87	0.81			0.80	0.81			
Factor 2	0.83	0.76	0.43	0.76		0.80	0.76			0.52	0.76			
Factor 3	0.77	0.72	0.75	0.72		0.82	0.72			0.74	0.72			

Note. [a] = The full labels of all items used in this analysis and their correspondence to items labels reported in this Table are fully disclosed in the online supplements; Non-significant parameters ($p \leq 0.05$) are italicized; Main a priori factor loadings are bolded; ICM= Independent cluster model; CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modelling; λ = Standardized factor loading; δ = Standardized uniqueness; G-Factor: Global factor from a bifactor model; S-Factor: Specific factor from a bifactor model; ω = Omega coefficient of composite reliability; α = Alpha coefficient of composite reliability.

Table 3. Relationships between CSM Factors and Body Mass Index (BMI)

	Relationships with BMI [β (s.e.)] in the various models:			
Factor	**ICM-CFA**	**ESEM**	**B-CFA**	**B-ESEM**
G-Factor (Diurnal Preference)	—	—	0.02 (0.03)	-0.01 (0.04)
Factor 1 (Morning Affect)	**-0.21 (0.05)**	**-0.16 (0.04)**	**-0.10 (0.04)**	**-0.09 (0.03)**
Factor 2 (Time of Rising)	**0.30 (0.08)**	**0.20 (0.05)**	**0.10 (0.05)**	**0.15 (0.04)**
Factor 3 (Bedtime Preference)	**-0.14 (0.06)**	**-0.07 (0.04)**	-0.03 (0.03)	-0.05 (0.04)
BMI R^2	0.03	0.03	0.02	0.03

Notes: Standardized regression coefficients (β) are reported, with standard errors (s.e.) in parentheses, significant differences are in bold ($p \leq 0.05$); ICM= Independent cluster model; CFA = Confirmatory factor analysis; B = Bifactor model; ESEM = Exploratory structural equation modelling; G-Factor: Global factor from a bifactor model; In bifactor models (B-CFA and B-ESEM), "Factors" are in fact S-factors (Specific factors); BMI = Body Mass Index; R^2 = Proportion of explained variance.

Table 4. Latent Means Comparison Between Groups Formed on the Basis of Gender and Age.

Latent variables	**Younger Females**	**Younger Males**	**Older Females**	**Older Males**
G-Factor (Diurnal Preference)	.00	.12 (.13)	.14 (.09)	.18 (.11)
S-Factor 1 (Morning Affect)	.00	-.04 (.14)	**.22 (.10)**	.15 (.11)
S-Factor 2 (Time of Rising)	.00	**.45 (.17)**	-.00 (.12)	**.57 (.13)**
S-Factor 3 (Bedtime Preference)	.00	**-.94 (.14)**	.05 (.11)	**-.57 (.12)**
G-Factor (Diurnal Preference)	-.12 (.13)	.00	.02 (.13)	.06 (.11)
S-Factor 1 (Morning Affect)	.04 (.14)	.00	.27 (.14)	.19 (.13)
S-Factor 2 (Time of Rising)	**-.45 (.17)**	.00	**-.45 (.17)**	.12 (.14)
S-Factor 3 (Bedtime Preference)	**.94 (.14)**	.00	**.98 (.14)**	**.37 (.13)**
G-Factor (Diurnal Preference)	-.14 (.09)	-.02 (.13)	.00	.04 (.10)
S-Factor 1 (Morning Affect)	**-.22 (.10)**	-.27 (.14)	.00	-.07 (.11)
S-Factor 2 (Time of Rising)	.01 (.12)	**.46 (.17)**	.00	**.58 (.13)**
S-Factor 3 (Bedtime Preference)	-.05 (.11)	**-.98 (.14)**	.00	**-.61 (.12)**
G-Factor (Diurnal Preference)	-.18 (.11)	-.06 (.11)	-.04 (.10)	.00
S-Factor 1 (Morning Affect)	-.15 (.11)	-.19 (.13)	.07 (.11)	.00
S-Factor 2 (Time of Rising)	**-.58 (.13)**	-.12 (.14)	**-.58 (.13)**	.00
S-Factor 3 (Bedtime Preference)	**.57 (.12)**	**-.37 (.13)**	**.61 (.12)**	.00

Notes: Latent means are reported, with standard errors in parentheses, significant differences are in bold ($p \leq 0.05$); In this table, the latent means are fixed to zero in one referent group for identification purposes, and the latent means (and their significance) estimated in the other groups reflect deviations from this referent groups expressed in standard deviation units; G-Factor: Global factor from a bifactor model; S-Factor: Specific factor from a bifactor model.

Online Supplements for:

Exploring Sources of Construct-Relevant Multidimensionality in Psychiatric Measurement: A Tutorial and Illustration using the Composite Scale of Morningness

These online supplements are to be posted on the journal website and hot-linked to the manuscript. If the journal does not offer this possibility, these materials can alternatively be posted on one of our personal websites (we will adjust the in-text reference upon acceptance).

Some Theoretical Background on the Assessment of Diurnal Preference

Chronotype or diurnal preference is an important inter-individual difference related to the time of day when a person is the most alert and awake, and to preferences for early or late awakening. Many instruments exist to assess diurnal preferences, including the Morningness-Eveningness Questionnaire (MEQ) (Horne and Östberg, 1976), the Diurnal-Type scale (DTS) (Torsvall and Âkerstedt, 1980) and the Circadian-Type Questionnaire (CTQ) (Folkard et al., 1979). The Composite Scale of Morningness (CSM) was created through the selection of the "best", or most discriminant items available in the MEQ, DTS and CTQ (Smith et al., 1989). The authors used principal component analyses (PCA) to reduce a pool of 26 items to 19 items. Then, they discarded the only item that loaded onto a fourth component, three items with low item-total correlations, and three items with content that did not match the component they are assumed to reflect. This procedure led to a 13-item scale assessing three components: *Morning Activities* (items 1,6,8,9,10,11,13), *Morning Affect* (items 3,4,5,12) and *Evening Activities* (items 2,7) (see next section for the items).

Since then, researchers interested in the structure of the CSM have found results supporting solutions including 1 to 3 factors (Di Milia et al., 2013). Unfortunately, previous EFA studies (Caci et al., 2005; Diaz-Morales and Sanchez-Lopez, 2005; Gil et al., 2008; Önder et al., 2013) have tended to rely on suboptimal criteria to select the number of factors (Preacher and MacCallum, 2003), while CFA studies have generally converged on solutions presenting only marginal fit to the data (Diaz Morales and Sanchez-Lopez, 2004; Randler and Diaz-Morales, 2007), forcing researchers to rely on suboptimal post-hoc modifications (Randler and Diaz-Morales, 2007). The *Morning Affect* factor consistently appears in all studies (Caci et al., 2005; Caci et al., 2009; Gil et al., 2008; Önder et al., 2013; Smith et al., 2002; Smith et al., 1989). However, the other factors emerge inconsistently, and with a changing content. Part of the reason for this could be that the initial study was based on PCA, which typically tends to extract a large first principal component explaining a maximum of variance, whereas reflective procedures (EFA/CFA) achieve a better distribution of the total covariance among factors. In support of this hypothesis, at least two of the items (9,13) included in the original *Morning Activities* component could equally be related to morningness or eveningness. Indeed, a recent study using the French CSM supported an alternative solution in which these two items where moved to the *Evening Activities* factor (Caci et al., 2005) – relabelled *Activity Planning*. This model thus represents a promising solution to the previous uncertainty regarding the CSM structure.

We first contrast ICM-CFA, B-CFA, ESEM, and B-ESEM representations of responses to the CSM. Then, we conduct predictive analyses to estimate the relations between CSM factors and Body Mass Index [BMI: weight(kg)/height(m)^2]. Recent studies suggested positive relations between eveningness and obesity (Wang, 2014), as well as with BMI and unhealthy eating habits in obese patients (Lucassen et al., 2013). The current study thus tests whether these results extend to a more general population sample. We also conduct tests of measurement invariances across subgroups of participants formed on the basis of age, gender, and combinations. Tests of measurement invariances are an important pre-requisite to unbiased group comparisons (Meredith, 1993). As such, these tests verify the extent to which the CSM factor structure generalizes across males and females of different age groups, which is interesting in light of reports that diurnal preferences may present variations among gender and age groups (Caci et al., 2005; Kim et al., 2002; Smith et al., 2002).

Composite Scale of Morningness

Smith CS, Reilly C, Midkiff K. Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. J Appl Psychol. 1989;74(5):728-38.

Items 3, 4, 5, 11 reflect a preference for mornings, whereas items 1, 2, 6, 7, 8, 9, 10, 12, 13 reflect a preference for evenings.

Please check the response for *each* item that best describes *you*.

1. Considering only your own "feeling best" rhythm, at what time would you get up if you were entirely free to plan you day?
   - 5:00-6:30 a.m.
   - 6:30-7:45 a.m.
   - 7:45-9:45 a.m.
   - 9:45-11:00 a.m.
   - 11:00 a.m.-12:00 (noon)
2. Considering only your "feeling best" rhythm, at what time would you go to bed if you were entirely free to plan your evening?
   - 8:00-9:00 p.m.
   - 9:00-10:15 p.m.
   - 10:15 p.m.-12:30 a.m.
   - 12:30 a.m.-1:45 a.m.
   - 1:45 a.m.-3:00 a.m.
3. Assuming normal circumstances, how easy do you find getting up in the morning?
   - Not at all easy
   - Slight easy
   - Fairly easy
   - Very easy
4. How alert do you feel during the first half hour after having awakened in the morning?
   - Not at all alert
   - Slightly alert
   - Fairly alert
   - Very alert
5. During the first half hour after having awakened in the morning, how tired do you feel?
   - Very tired
   - Fairly tired
   - Fairly refreshed
   - Very refreshed
6. You have decided to engage in some physical exercise. A friend suggests that you do this one hour twice a week and the best time for him is 7:00-8:00 a.m. Bearing in mind nothing else but your own "feeling best" rhythm, how do you think you would perform?
   - Would be in good form
   - Would be in a reasonable form
   - Would find it difficult
   - Would find it very difficult
7. At what time in the evening do you feel tired and, as a result, in need of sleep?
   - 8:00-9:00 p.m.
   - 9:00-10:15 p.m.
   - 10:15 p.m.-12:30 a.m.
   - 12:30 a.m.-1:45 a.m.
   - 1:45 a.m.-3:00 a.m.

8. You wish to be at your peak performance for a test which you know is going to be mentally exhausting and lasting for two hours. You are entirely free to plan your day, and considering only you own "feeling best" rhythm, which ONE of the four testing times would you choose?
    - 8:00-10:00 a.m.
    - 11:00 a.m. - 1:00 p.m.
    - 3:00-5:00 p.m.
    - 7:00-9:00 p.m.
9. One hears about "morning" and "evening" types of people. Which ONE of these types do you consider yourself to be?
    - Definitively a morning type
    - More a morning than an evening type
    - More an evening than a morning type
    - Definitively an evening type
10. When would you prefer to rise (provided you have a full day's work - 8 hours) if you were totally free to arrange your time?
    - Before 6:30 a.m.
    - 6:30-7:30 a.m.
    - 7:30-8:30 a.m.
    - 8:30 a.m. or later
11. If you always had to rise at 6:00 a.m., what do you think it would be like?
    - Very difficult and unpleasant
    - Rather difficult and unpleasant
    - A little unpleasant but no great problem
    - Easy and not unpleasant
12. How long a time does it usually take before you "recover your senses" in the morning after rising from a night's sleep?
    - 0-10 minutes
    - 11-20 minutes
    - 21-40 minutes
    - more than 40 minutes
13. Please indicate to what extent you are a morning or evening *active* individual.
    - Pronounced morning active (morning alert and evening tired)
    - To some extent, morning active
    - To some extent, evening active
    - Pronounced evening active (morning tired and evening alert)

Discussion of the Substantive Results About the CSM

Our results also have relevance for our understanding of the CSM as an instrument and for extending knowledge on the construct of diurnal preference. First, our results supported the revised three factor-structure of the CSM (Caci et al., 2005). However, they also showed that the CSM items could be used to reflect a well-defined global diurnal preference factor. Over and above this global factor, most CSM items also serve to define specific factors related to *Morning Affect* and *Bedtime Preference*. An additional S-Factor related to *Time of Rising* retained a lower level of specificity but was shown to present meaningful associations with categorical (age-groups, gender) and continuous (BMI) covariates. More precisely, although the B-ESEM representation of CSM proved to be completely invariant across gender and age-groups, the results showed meaningful latent means differences between groups: (a) older females presented a more positive morning affect than younger females; (b) males preferred getting up earlier than females; (c) men preferred getting into bed later than females, and older males preferred doing so earlier than younger males. These results are generally in line with previous results (Caci et al., 2005; Carrier et al., 1997; Smith et al., 2002), and future studies are needed to further explore the reasons underlying these differences. Finally, and perhaps most interestingly, the results showed small but meaningful relations between *Morning Affect* and lower levels of BMI, and between *Time of Rising* and higher levels of BMI. These results are particularly important as they extend the results from previous studies (Lucassen et al., 2013) suggesting relations between diurnal preference and obesity. Our results show that this relation generalizes to more normative BMI variations, but also differs according to the specific dimension of diurnal preference that is considered. Future research should devote more attention to the mechanisms underlying these relations.

References used in this Appendix but not in the main body of text

Caci H, Adan A, Bohlé P, Natale V, Pornpitakpan C, Tilley A. (2005). Transcultural properties of the Composite Scale of Morningness: the relevance of the 'Morning Affect' factor. *Chronobiology International* **22**(3): 523-540.

Caci H, Deschaux O, Adan A, Natale V. (2009). Comparing three morningness scales: age and gender effects, structure and cut-off criteria. *Sleep Medicine* **10**(2): 240-245.

Carrier J, Monk TH, Buysse DJ, Kupfer DJ. (1997). Sleep and morningness-eveningness in the middle years of life (20-59 years). *J Sleep Res* **6**(4): 230-237.

Di Milia L, Adan A, Natale V, Randler C. (2013). Reviewing the psychometric properties of contemporary circadian typology measures. *Chronobiology International* **30**(10): 1261-1271.

Diaz Morales JF, Sanchez-Lopez MP. (2004). Composite and Preferences Scales of Morningness: reliability and factor invariance in adult and university samples. *The Spanish Journal of Psychology* **7**(2): 93-100.

Diaz-Morales JF, Sanchez-Lopez MP. (2005). Composite scales of morningness and preferences: preliminary validity data in Peruvian undergraduates. *Ergonomics* **48**(4): 354-363.

Folkard S, Monk T, Lobban M. (1979). Towards a predictive test of adjustment to shift work. *Ergonom* **22**: 79-91.

Gil E, Abdo PL, Rodriguez M, Zanin L, De Bortoli M. (2008). Psychometric and comparative study of an Argentine version of the morningness composite and the early/late preference scales. *Chronobiology International* **25**(1): 133-143.

Horne JA, Östberg O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Internat J Chronobiol* **4**: 97-110.

Kim S, Dueker GL, Hasher L, Goldstein D. (2002). Children's time of day preference: age, gender and ethnic differences. *Personality and Individual Differences* **33**(7): 1083-1090.

Lucassen EA, Zhao X, Rother KI, Mattingly MS, Courville AB, de Jonge L, Csako G, Cizza G, Group ftSES. (2013). Evening chronotype is associated with changes in eating behavior, more sleep apnea, and increased stress hormones in short sleeping obese individuals. *PLoS One* **8**(3): e56519, DOI: 10.1371/journal.pone.0056519.

Meredith W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* **58**(4): 525-543.

Önder I, Besoluk S, Baris Horzum M. (2013). Psychometric properties of the Turkish version of the Composite Scale of Morningness. *Spanish Journal of Psychology* **16**(e67): 1-10.

Preacher KJ, MacCallum RC. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics* **2**(1): 13-43.

Randler C, Diaz-Morales JF. (2007). Morningness in German and Spanish students: a comparative study. *European Journal of Personality* **21**(4): 419-427.

Smith CS, Folkard S, Schmieder RA, Parra LF, Spelten E, Almiral H, Sen RN, Sahu S, Perez LM, Tisak J. (2002). Multinational investigation of morning-evening orientation using a new measure of morningness. *Personality and Individual Differences* **32**(6): 949-968.

Smith CS, Reilly C, Midkiff K. (1989). Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. *J Appl Psychol* **74**(5): 728-738.

Torsvall L, Âkerstedt T. (1980). A diurnal type scale. Construction, consistency and validation in shift work. *Scandinavian Journal of Work, Environment and Health* **6**: 283-290.

Wang L. (2014). Body mass index, obesity, and self-control: a comparison of chronotypes. *Social Behavior and Personality: an international journal* **42**(2): 313-320.

Title: ICM-CFA

```
! Statements preceded by ! are comments not part of the input setup.
! The following statement is used to identify the data file. Here, the data file is labelled CSM.dat.
DATA:
FILE IS CSM.dat;
! The variables names function identifies all variables in the data set, in order of appearance.
! The usevariables command identifies the variables used in the analysis.
! The categorical command identifies the variables that are ordered-categorical in nature.
VARIABLE:
NAMES ARE SEX AGE BMI CAT CSM1-CSM13;
USEVARIABLES ARE CSM1-CSM13;
CATEGORICAL ARE CSM1-CSM13;
! The next section defines the analysis. Here robust weighed least square estimation
! (WLSMV) is used.
! With WLSMV estimation, it is often useful to increase the number of iterations.
ANALYSIS:
ESTIMATOR IS WLSMV;
ITERATIONS = 10000;
! The next section defines the model.
! The @ symbol is use to fix parameter estimates to a specific value.
! The * symbol indicates the free estimation of a parameter value (or provide a start value).
! Each input lines ends with ;
! Factor loadings are noted with BY, regressions with ON, correlations with WITH,
! means and thresholds are noted between brackets [];
! variances and residuals are noted without brackets.
! Here, An ICM-CFA model is specified
! with 3 factors (F1 to F3) defined by their respective items.
! the model is identified by fixing the factor variance to 1 (F1-F3@1), allowing first factor
loading to be freely estimated (*)
MODEL:
 F1 BY CSM3* CSM4 CSM5 CSM12;
 F2 BY CSM1* CSM6 CSM8 CSM10 CSM11;
 F3 BY CSM2* CSM7 CSM9 CSM13;
 F1@1; F2@1; F3@1;
 F1 WITH F2 F3;
 F2 WITH F3;
! Specific sections of output are requested.
OUTPUT:
  SAMPSTAT STANDARDIZED RESIDUAL CINTERVAL TECH1 TECH2  TECH4
MODINDICES(ALL);
```

! Redundant sections are not repeated, we only focus on sections that differ from previous models.

Title: Bifactor-CFA

! A bifactor CFA model is specified with 3 specific factors (FS1 to FS3)
! All items are also used to define a global factor FG.
! All factors are set to be orthogonal (correlations @0)

```
MODEL:
 FG BY  CSM4* CSM3 CSM5 CSM12 CSM1 CSM6 CSM8 CSM10 CSM11 CSM2 CSM7 CSM9
 CSM13;
 FS1 BY CSM3* CSM4 CSM5 CSM12;
 FS2 BY CSM1* CSM6 CSM8 CSM10 CSM11;
 FS3 BY CSM2* CSM7 CSM9 CSM13;
 F1@1; F2@1; F3@1;
 FG@1;
 FS1 WITH FS2 FS3@0;
 FS2 WITH FS3@0;
 FG WITH FS1 FS2 FS3@0;
```

Title: ESEM

```
! An ESEM model is specified with target oblique rotation.
ANALYSIS:
ESTIMATOR IS WLSMV; ITERATIONS = 10000;
ROTATION = TARGET;
! The factors (F1 to F3) are defined with main loadings from their respective items,
! In addition to these main loadings, all other cross-loadings are estimated but targeted
! to be as close to 0 as possible (~0).
! Factors forming a single set of ESEM factors (with cross-loadings between factors)
! are indicated by using the same label in parenthesis after * (*1).
MODEL:
F1 BY CSM1~0 CSM2~0 CSM3 CSM4 CSM5 CSM6~0
CSM7~0 CSM8~0 CSM9~0 CSM10~0 CSM11~0 CSM12 CSM13~0 (*1);
F2 BY CSM1 CSM2~0 CSM3~0 CSM4~0 CSM5~0
CSM6 CSM7~0 CSM8 CSM9~0 CSM10 CSM11 CSM12~0 CSM13~0 (*1);
F3 BY CSM1~0 CSM2 CSM3~0 CSM4~0 CSM5~0
CSM6~0 CSM7 CSM8~0 CSM9 CSM10~0 CSM11~0 CSM12~0 CSM13 (*1);
```

```
Title: Bifactor-ESEM
! A Bifactor-ESEM model is specified with orthogonal target rotation.
ANALYSIS:
ESTIMATOR IS WLSMV; ITERATIONS = 10000;
ROTATION = TARGET (orthogonal);
! The specific factors (FS1 to FS3) are defined with main loadings from their respective items.
! All other cross-loadings are estimated but targeted to be as close to 0 as possible (~0).
! The global factor is defined through main loadings from all items, and is included in
! the same set of ESEM factors as FS1-FS3 (*1)
MODEL:
FG BY CSM1 CSM2 CSM3 CSM4 CSM5 CSM6
CSM7 CSM8 CSM9 CSM10 CSM11 CSM12 CSM13 (*1);

FS1 BY CSM1~0 CSM2~0 CSM3 CSM4 CSM5 CSM6~0
CSM7~0 CSM8~0 CSM9~0 CSM10~0 CSM11~0 CSM12 CSM13~0 (*1);
FS2 BY CSM1 CSM2~0 CSM3~0 CSM4~0 CSM5~0
CSM6 CSM7~0 CSM8 CSM9~0 CSM10 CSM11 CSM12~0 CSM13~0 (*1);
FS3 BY CSM1~0 CSM2 CSM3~0 CSM4~0 CSM5~0
CSM6~0 CSM7 CSM8~0 CSM9 CSM10~0 CSM11~0 CSM12~0 CSM13 (*1);
```

Title: Including an Outcome Variable
! The predictor (here BMI), is added to the usevariables list.
! Because BMI is a continuous variable, it is not added to the categorical list.
VARIABLE:
NAMES ARE SEX AGE BMI CAT CSM1-CSM13;
USEVARIABLES ARE CSM1-CSM13;
CATEGORICAL ARE CSM1-CSM13;
! Then, in the model sections, the following statements are added to indicate that the factors are used
! to predict BMI:
! ICM-CFA and ESEM:
IMC ON F1 F2 F3;
! Bifactor-CFA and Bifactor-ESEM:
IMC ON FG FS1 FS2 FS3;

Title: Measurement Invariance across Gender – Configural Invariance.

```
DATA:
FILE IS CSM.dat;
! The grouping variable is used to identify the groups and labels are given to each of the value
! An observed grouping variables does not need to be included in the usevariables or categorical list.
VARIABLE:
NAMES ARE SEX AGE BMI CAT CSM1-CSM13;
USEVARIABLES ARE CSM1-CSM13;
CATEGORICAL ARE CSM1-CSM13;
GROUPING IS SEX (1=women 2=men);
! As before.
! Parameterization = theta is added in order to be able to test for the invariance of uniquenesses.
ANALYSIS:
  TYPE IS GENERAL;
  ESTIMATOR IS WLSMV;
  ITERATIONS = 10000;
  PARAMETERIZATION=THETA;
  ROTATION = TARGET (orthogonal);
! The global model section is used to define the global model used in both groups.
! Parameter freely estimated across groups will be specified in group-specific sections.
! See Morin, Moullec et al. (2011) and Guay, Morin et al. (2014), cited in main manuscript, for
!additional details on specifications of invariance testing for WLSMV estimation.
! The first part is as above for the bifactor-ESEM model.
MODEL:
FG BY CSM1 CSM2 CSM3 CSM4 CSM5 CSM6
CSM7 CSM8 CSM9 CSM10 CSM11 CSM12 CSM13 (*1);
FS1 BY CSM1~0 CSM2~0 CSM3 CSM4 CSM5 CSM6~0
CSM7~0 CSM8~0 CSM9~0 CSM10~0 CSM11~0 CSM12 CSM13~0 (*1);
FS2 BY CSM1 CSM2~0 CSM3~0 CSM4~0 CSM5~0
CSM6 CSM7~0 CSM8 CSM9~0 CSM10 CSM11 CSM12~0 CSM13~0 (*1);
FS3 BY CSM1~0 CSM2 CSM3~0 CSM4~0 CSM5~0
CSM6~0 CSM7 CSM8~0 CSM9 CSM10~0 CSM11~0 CSM12~0 CSM13 (*1);
! Here, thresholds are specified. For X answers categories, there are X-1 thresholds to be specified
! using the $1, $1, $3, etc. symbol. By default, thresholds are invariant across groups so only non-
! invariant thresholds need to be specified here.
! For tests of configural invariance, the first threshold from each item is set to be invariant, and the
! second threshold from a referent indicator for each factor.
!Referent indicator for FS2 (5 categories):
[CSM1$3]; [CSM1$4];
!Referent indicator for FS3 (5 categories):
[CSM2$3]; [CSM2$4];
!Referent indicator for FS1 (4 categories):
 [CSM3$3];
!Referent indicator for FG (4 categories):
 [CSM4$3];
! Other indicators:
[CSM5$2]; [CSM5$3];
[CSM6$2]; [CSM6$3];
[CSM7$2]; [CSM7$3]; [CSM7$4];
[CSM8$2]; [CSM8$3];
[CSM9$2]; [CSM9$3];
[CSM10$2]; [CSM10$3];
[CSM11$2]; [CSM11$3];
[CSM12$2]; [CSM12$3];
[CSM13$2]; [CSM13$3];
```

```
! In the group specific statement, all parameter to be freely estimated in the other group are specified.
! For X groups, X-1 group-specific statement are needed, starting with group 2.
! By default, the factor variances are set to 1 in all groups
! which is as should be for configural invariance.
! By default the latent means are fixed to 0 in group 1, and freely estimated in group 2 and subsequent,
! which is as should be for configural invariance.
! By default, uniquenesses are set to 1 in the first group and freely estimated in the other groups,
! which is as should be for configural invariance.
MODEL men:
FG BY CSM1 CSM2 CSM3 CSM4 CSM5 CSM6
CSM7 CSM8 CSM9 CSM10 CSM11 CSM12 CSM13 (*1);
FS1 BY CSM1~0 CSM2~0 CSM3 CSM4 CSM5 CSM6~0
CSM7~0 CSM8~0 CSM9~0 CSM10~0 CSM11~0 CSM12 CSM13~0 (*1);
FS2 BY CSM1 CSM2~0 CSM3~0 CSM4~0 CSM5~0
CSM6 CSM7~0 CSM8 CSM9~0 CSM10 CSM11 CSM12~0 CSM13~0 (*1);
FS3 BY CSM1~0 CSM2 CSM3~0 CSM4~0 CSM5~0
CSM6~0 CSM7 CSM8~0 CSM9 CSM10~0 CSM11~0 CSM12~0 CSM13 (*1);
!Referent indicator for FS2 (5 categories):
[CSM1$3]; [CSM1$4];
!Referent indicator for FS3 (5 categories):
[CSM2$3]; [CSM2$4];
!Referent indicator for FS1 (4 categories):
 [CSM3$3];
!Referent indicator for FG (4 categories):
 [CSM4$3];
! Other indicators:
[CSM5$2]; [CSM5$3];
[CSM6$2]; [CSM6$3];
[CSM7$2]; [CSM7$3]; [CSM7$4];
[CSM8$2]; [CSM8$3];
[CSM9$2]; [CSM9$3];
[CSM10$2]; [CSM10$3];
[CSM11$2]; [CSM11$3];
[CSM12$2]; [CSM12$3];
[CSM13$2]; [CSM13$3];
! Specific sections of output are requested.
OUTPUT:
  SAMPSTAT STANDARDIZED RESIDUAL CINTERVAL TECH1 TECH2  TECH4
MODINDICES(ALL);
! The following section is used to request a save data file to be used in the calculation of Chi-
! square differences tests based on WLSMV estimation.
SAVEDATA:
  DIFFTEST = BESEM_sex_conf.dat;
```

```
! Redundant sections are not repeated, we only focus on sections that differ from previous models.
Title: Measurement Invariance across Gender – Weak Invariance.
! The DIFFTEST function is used to request a chi square difference test, using the saved data file from
! the previous model in the sequence.
ANALYSIS:
  TYPE IS GENERAL;
  ESTIMATOR IS WLSMV;
  ITERATIONS = 10000;
  PARAMETERIZATION=THETA;
  ROTATION = TARGET (orthogonal);
  DIFFTEST = BESEM_sex_conf.dat;
! The only difference between this model and the previous one is that the specification of factor
! loadings is not repeated in the group-specific section. These are invariant at default. Also, when the
! loadings are invariant, the factor variances are freely estimated in all groups but the first.
MODEL:
FG BY CSM1 CSM2 CSM3 CSM4 CSM5 CSM6
CSM7 CSM8 CSM9 CSM10 CSM11 CSM12 CSM13 (*1);
FS1 BY CSM1~0 CSM2~0 CSM3 CSM4 CSM5 CSM6~0
CSM7~0 CSM8~0 CSM9~0 CSM10~0 CSM11~0 CSM12 CSM13~0 (*1);
FS2 BY CSM1 CSM2~0 CSM3~0 CSM4~0 CSM5~0
CSM6 CSM7~0 CSM8 CSM9~0 CSM10 CSM11 CSM12~0 CSM13~0 (*1);
FS3 BY CSM1~0 CSM2 CSM3~0 CSM4~0 CSM5~0
CSM6~0 CSM7 CSM8~0 CSM9 CSM10~0 CSM11~0 CSM12~0 CSM13 (*1);
 [CSM1$3]; [CSM1$4];
 [CSM2$3]; [CSM2$4];
 [CSM3$3];
 [CSM4$3];
 [CSM5$2]; [CSM5$3];
[CSM6$2]; [CSM6$3];
[CSM7$2]; [CSM7$3]; [CSM7$4];
[CSM8$2]; [CSM8$3];
[CSM9$2]; [CSM9$3];
[CSM10$2]; [CSM10$3];
[CSM11$2]; [CSM11$3];
[CSM12$2]; [CSM12$3];
[CSM13$2]; [CSM13$3];
MODEL men:
 [CSM1$3]; [CSM1$4];
 [CSM2$3]; [CSM2$4];
 [CSM3$3];
 [CSM4$3];
 [CSM5$2]; [CSM5$3];
[CSM6$2]; [CSM6$3];
[CSM7$2]; [CSM7$3]; [CSM7$4];
[CSM8$2]; [CSM8$3];
[CSM9$2]; [CSM9$3];
[CSM10$2]; [CSM10$3];
[CSM11$2]; [CSM11$3];
[CSM12$2]; [CSM12$3];
[CSM13$2]; [CSM13$3];
SAVEDATA:
  DIFFTEST = BESEM_sex_weak.dat;
```

Title: Measurement Invariance across Gender – Strong Invariance.

```
ANALYSIS:
  TYPE IS GENERAL;
  ESTIMATOR IS WLSMV;
  ITERATIONS = 10000;
  PARAMETERIZATION=THETA;
  ROTATION = TARGET (orthogonal);
  DIFFTEST = BESEM_sex_weak.dat;
! The only difference between this model and the previous one is that thresholds are invariants across
! group by default and thus do not need to be specified.
! When thresholds are invariant, the factor means are freely estimated in all groups but the first.
! Again: Loadings are invariant by default; variances and uniquenesses are fixed to be 1 in the first
! group and free in the other groups.
MODEL:
FG BY CSM1 CSM2 CSM3 CSM4 CSM5 CSM6
CSM7 CSM8 CSM9 CSM10 CSM11 CSM12 CSM13 (*1);
FS1 BY CSM1~0 CSM2~0 CSM3 CSM4 CSM5 CSM6~0
CSM7~0 CSM8~0 CSM9~0 CSM10~0 CSM11~0 CSM12 CSM13~0 (*1);
FS2 BY CSM1 CSM2~0 CSM3~0 CSM4~0 CSM5~0
CSM6 CSM7~0 CSM8 CSM9~0 CSM10 CSM11 CSM12~0 CSM13~0 (*1);
FS3 BY CSM1~0 CSM2 CSM3~0 CSM4~0 CSM5~0
CSM6~0 CSM7 CSM8~0 CSM9 CSM10~0 CSM11~0 CSM12~0 CSM13 (*1);
MODEL men:
! Empty
SAVEDATA:
  DIFFTEST = BESEM_sex_strong.dat;
```

Title: Measurement Invariance across Gender – Strict Invariance.

```
ANALYSIS:
  TYPE IS GENERAL;
  ESTIMATOR IS WLSMV;
  ITERATIONS = 10000;
  PARAMETERIZATION=THETA;
  ROTATION = TARGET (orthogonal);
  DIFFTEST = BESEM_sex_strong.dat;
! The only difference between this model and the previous one is that here uniquenesses are set to be
! fixed to 1 in all groups.
! Again: Loadings and thresholds are invariant by default;
! Variances and uniquenesses are fixed to be 1 in the first group and free in the other groups.
! Means are fixed to be 0 in the first group and free in the other groups.
MODEL:
FG BY CSM1 CSM2 CSM3 CSM4 CSM5 CSM6
CSM7 CSM8 CSM9 CSM10 CSM11 CSM12 CSM13 (*1);
FS1 BY CSM1~0 CSM2~0 CSM3 CSM4 CSM5 CSM6~0
CSM7~0 CSM8~0 CSM9~0 CSM10~0 CSM11~0 CSM12 CSM13~0 (*1);
FS2 BY CSM1 CSM2~0 CSM3~0 CSM4~0 CSM5~0
CSM6 CSM7~0 CSM8 CSM9~0 CSM10 CSM11 CSM12~0 CSM13~0 (*1);
FS3 BY CSM1~0 CSM2 CSM3~0 CSM4~0 CSM5~0
CSM6~0 CSM7 CSM8~0 CSM9 CSM10~0 CSM11~0 CSM12~0 CSM13 (*1);
MODEL men:
CSM1-CSM13@1;
SAVEDATA:
  DIFFTEST = BESEM_sex_strict.dat;
```

Title: Measurement Invariance across Gender – Variance-Covariance Invariance.

```
ANALYSIS:
  TYPE IS GENERAL;
  ESTIMATOR IS WLSMV;
  ITERATIONS = 10000;
  PARAMETERIZATION=THETA;
  ROTATION = TARGET (orthogonal);
  DIFFTEST = BESEM_sex_strict.dat;
! Here, the variances are fixed to 1 in all groups.
MODEL:
FG BY CSM1 CSM2 CSM3 CSM4 CSM5 CSM6
CSM7 CSM8 CSM9 CSM10 CSM11 CSM12 CSM13 (*1);
FS1 BY CSM1~0 CSM2~0 CSM3 CSM4 CSM5 CSM6~0
CSM7~0 CSM8~0 CSM9~0 CSM10~0 CSM11~0 CSM12 CSM13~0 (*1);
FS2 BY CSM1 CSM2~0 CSM3~0 CSM4~0 CSM5~0
CSM6 CSM7~0 CSM8 CSM9~0 CSM10 CSM11 CSM12~0 CSM13~0 (*1);
FS3 BY CSM1~0 CSM2 CSM3~0 CSM4~0 CSM5~0
CSM6~0 CSM7 CSM8~0 CSM9 CSM10~0 CSM11~0 CSM12~0 CSM13 (*1);
! The unrotated covariances (even if the rotated covariances are orthogonal) need to be fixed to
! invariance across groups. The labels in parentheses indicate that these covariance are fixed
! to invariance across groups.
FS1 WITH FS2 (c1);
FS1 WITH FS3 (c2);
FS2 WITH FS3 (c3);
FG WITH FS1 (c4);
FG WITH FS2 (c5);
FG WITH FS3 (c6);
MODEL men:
CSM1-CSM13@1;
FG@1; FS1@1; FS2@1; FS3@1;
SAVEDATA:
  DIFFTEST = BESEM_sex_vc.dat;
```

Title: Measurement Invariance across Gender – Latent Mean Invariance.

```
ANALYSIS:
  TYPE IS GENERAL;
  ESTIMATOR IS WLSMV;
  ITERATIONS = 10000;
  PARAMETERIZATION=THETA;
  ROTATION = TARGET (orthogonal);
  DIFFTEST = BESEM_sex_vc.dat;
! Here, the means are fixed to 0 in all groups.
MODEL:
FG BY CSM1 CSM2 CSM3 CSM4 CSM5 CSM6
CSM7 CSM8 CSM9 CSM10 CSM11 CSM12 CSM13 (*1);
FS1 BY CSM1~0 CSM2~0 CSM3 CSM4 CSM5 CSM6~0
CSM7~0 CSM8~0 CSM9~0 CSM10~0 CSM11~0 CSM12 CSM13~0 (*1);
FS2 BY CSM1 CSM2~0 CSM3~0 CSM4~0 CSM5~0
CSM6 CSM7~0 CSM8 CSM9~0 CSM10 CSM11 CSM12~0 CSM13~0 (*1);
FS3 BY CSM1~0 CSM2 CSM3~0 CSM4~0 CSM5~0
CSM6~0 CSM7 CSM8~0 CSM9 CSM10~0 CSM11~0 CSM12~0 CSM13 (*1);
FS1 WITH FS2 (c1);
FS1 WITH FS3 (c2);
FS2 WITH FS3 (c3);
FG WITH FS1 (c4);
FG WITH FS2 (c5);
FG WITH FS3 (c6);
MODEL men:
CSM1-CSM13@1;
FG@1; FS1@1; FS2@1; FS3@1;
[FG@0]; [F1@0]; [F2@0]; [F3@0];
```

Title: Measurement Invariance across Age
! Here we used age-groups. In the data set, age is measured as a continuous variable, so it needs to be
! recoded. This is achieved using the cut command of the define function. The number in parentheses
! indicates the maximum value of age included in the category 0. Every value over that is coded 1.
! Then the grouping function of the variable command defines and labels the two groups thus created.
! The rest of the code for tests of invariance is then exactly as in the previous examples using gender
! as the grouping variable, except that the group sections of the model commands are defined using
! the new labels.
VARIABLE:
NAMES ARE SEX AGE BMI CAT CSM1-CSM13;
USEVARIABLES ARE CSM1-CSM13;
CATEGORICAL ARE CSM1-CSM13;
GROUPING IS age (0=younger 1=older);
DEFINE:
CUT AGE (39.999999);
[…]
Model:
[…]
Model older:
[…]

Title: Measurement Invariance across Age * Gender

! Here the If function of the define command is used to define four groups based on the combination o
! information from the sex variable defining gender and the continuous age variable. The function EQ
! means "equal" (e.g. sex EQ 1 identifies women); the function LT means "lower than" (LT 40
! identifies participants aged less than 40 exclusively); the function GE means "greater or equal to"
! (GE 40 identifies participants aged 40 inclusively or more).
! The four groups are then identified and labeled using the grouping function (YF: younger females;
! OF: older females; YM: younger males; OM: older males).
! Variables created using the define function need to be added as the end of the usevariables list.
! The rest of the code for tests of invariance is then exactly as in the previous examples using gender
! as the grouping variable, except that there are now more group sections in the model, and these are
! defined using the new labels.

```
VARIABLE:
NAMES ARE SEX AGE BMI CAT CSM1-CSM13;
USEVARIABLES ARE CSM1-CSM13 group;
CATEGORICAL ARE CSM1-CSM13;
GROUPING IS group (1=YF 2=OF 3=YM 4=OM);
DEFINE:
IF (SEX EQ 1 AND AGE LT 40) THEN group=1;
IF (SEX EQ 1 AND AGE GE 40) THEN group=2;
IF (SEX EQ 2 AND AGE LT 40) THEN group=3;
IF (SEX EQ 2 AND AGE GE 40) THEN group=4;
[…]
Model:
[…]
Model OF:
[…]
Model YM:
[…]
Model OM:
[…]
```