International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 25(4): 277–288 (2016)
Published online 12 August 2015 in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/mpr.1485

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

ALEXANDRE J.S. MORIN,[1] A. KATRIN ARENS,[2] ANTOINE TRAN[3] & HERVÉ CACI[3]

1 Institute for Positive Psychology and Education, Australian Catholic University, Strathfield, NSW, Australia
2 German Institute for International Educational Research, Frankfurt, Germany
3 Hôpitaux Pédiatriques de Nice CHU Lenval, Nice, France

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

Correspondence
Alexandre J.S. Morin, Institute for Positive Psychology and Education, Australian Catholic University, Strathfield Campus, Locked Bag 2002, Strathfield, New South Wales 2135, Australia.
Email: alexandre.morin@acu.edu.au

Received 28 January 2015;
revised 11 May 2015;
accepted 19 May 2015

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: (1) assessment of conceptually adjacent constructs; (2) 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 1159 adults [655 female, 504 male, mean age (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. *Copyright © 2015 John Wiley & Sons, Ltd.*

Introduction

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 core component of ADHD should co-exist with more specific 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 Factor Analysis (EFA) or Confirmatory Factor Analysis (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, Jennrich 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

Int. J. Methods Psychiatr. Res. 25(4): 277–288 (2016). DOI: 10.1002/mpr

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.

Earlier, 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 B-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 0.100) 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 Analysis (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 (structural equation modelling) (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 lead to 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 and Maïano, 2011; 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 the two aforementioned 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 (the children and parent ADHD and related disorders 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 1000 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 (two to four 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 body mass index (BMI). The final sample used in this study thus includes 1159 parents [22–65 years old; mean age (M_{age}) = 41.84; mean body mass index (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 the 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 squares (WLSMV) estimator which outperforms Maximum Likelihood for ordered-categorical indicators with five 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 three-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 (males 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.*, 2015; 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); (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 0.900 and 0.950 for CFI and TLI, and lower than 0.080 and 0.060 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 0.010 or less and a RMSEA augmentation of 0.015 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 non-monotonic 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 < 0.900; RMSEA > 0.100). B-CFA and ESEM yield a clearly improved level of fit although they both remain marginal according to some indices (TLI < 0.950; RMSEA > 0.080). 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 (0.50–0.96; $M = 0.73$) and satisfactory composite reliability ($\omega = 0.77$–0.88). 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 (0.46–0.73; $M = 0.63$) 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 (0.30–0.57; $M = 0.45$) while all factors remain clearly defined (0.50–0.96; $M = 0.73$) and reliable ($\omega = 0.80$–0.87). However, although most cross-loadings remain small (|0.02–0.36|; $M = 0.14$), some are high enough (>0.30 for Items 3 and 9) to suggest that another source of unmodelled multidimensionality may be present, explaining the marginal fit of this model (TLI < 0.950; RMSEA > 0.080).

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 ($\lambda = 0.16$–0.77; $M = 0.59$; $\omega = 0.92$). Indeed, apart from two items associated with the *Activity Planning* S-factor that present a lower loading on the G-factor (item 2: $\lambda = 0.34$, "… at what time would you go to bed if you were entirely free to plan your evening?"; item 7: $\lambda = 0.16$, "At what time in the evening do you feel tired … ?"), the remaining items all have fully satisfactory loadings on the G-factor (0.50–0.77; $M = 0.65$) reflecting global diurnal preference. Interestingly, these two items present high loadings on their corresponding S-factor "Activity Planning" (0.61 and 0.85), whereas the two remaining indicators of this S-factor present much lower loadings (Item 9: $\lambda = 0.36$, "One hears about morning and evening

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; 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[1]	**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. [1]= 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.

types of people … ?"; Item 13: $\lambda = 0.36$, "… a morning or evening active individual?"). This suggests that these two items (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 ($\omega = 0.75$).

The results further reveal that the first S-factor (*Morning Affect*) retains a meaningful level of specificity ($\lambda = 0.37$–0.72; $M = 0.59$; $\omega = 0.78$), whereas the second S-factor (*Time of Rising*) apparently retains no meaningful specificity (|0.07–0.52|; $M = 0.23$; $\omega = 0.43$), 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 $<$ 0.950; RMSEA $>$ 0.080) and below the fit of ESEM (ΔCFI, ΔTLI, ΔRMSEA all $\geq$ 0.100), 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 ($\lambda = 0.50$–0.84; $M = 0.62$; $\omega = 0.92$) apart from Items 2 (0.29) and 7 (0.24) which together define a *Bedtime Preference* S-factor ($\lambda = 0.72$ and 0.73; versus 0.26 and 0.25 for Items 9 and 13; $\omega = 0.74$); (b) a well-defined *Morning Affect* S-factor (0.41–0.69; $M = 0.58$; $\omega = 0.80$); (c) a weakly defined *Time of Rising* S-factor (0.04–0.47; $M = 0.31$; $\omega = 0.52$). 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 (|0.01–0.23|; $M = 0.09$) than in ESEM (|0.02–0.36|; $M = 0.14$), 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 solutions, and in the B-CFA/B-ESEM solutions, 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

Table 3. Relationships between Composite Scale of Morningness (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 relations 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.

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

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.

multidimensionality naturally calls for bifactor models (Reise, 2012), whereas the second source rather calls for ESEM (Marsh *et al.*, 2014). Finally, B-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 B-ESEM framework of broad relevance to psychiatric, epidemiological and biomedical research. The implementation of this framework was illustrated while using a WLSMV estimator 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$0.10 or even $\geq$ 0.20) 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 B-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 all studies of any 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 B-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 B.O. (2006) Robust Chi-square Difference Testing with Mean and Variance Adjusted Test Statistics, Los Angeles, CA, Muthén and Muthén. http://www.statmodel.com/examples/webnote.shtml – web10

Beauducel A., Herzberg P.Y. (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 M.W. (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 A.J.S., 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 A.J.S., 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 F.F. (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, **14**(3), 464–504, DOI: 10.1080/10705510701301834

Chen F.F., West S.G., Sousa K.H. (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 G.W., Rensvold R.B. (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, **9**(2), 233–255, DOI: 10.1207/S15328007SEM0902_05

Clark S.L., Muthén B.O., Kaprio J., D'Onofrio B. M., Viken R., Rose R.J. (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 S.J., DiStefano C. (2006) Non-normal and categorical data in structural equation modeling. In Hancock G.R., Mueller R.O. (eds) Structural Equation Modeling: A Second Course, pp. 269–314, Greenwich, CT, IAP.

Gignac G.E., Watkins M.W. (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 A.J.S., Litalien D., Valois P., Vallerand R.J. (2015) An application of exploratory structural equation modeling to evaluate the Academic Motivation Scale. *The Journal of Experimental Education*, **83**(1), 51–82, DOI: 10.1080/00220973.2013.876231

Hu L.T., Bentler P.M. (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 R.I., Bentler P.M. (2011) Exploratory bi-factor analysis. *Psychometrika*, **76**(4), 537–549, DOI: 10.1007/s11336-011-9218.4

Marsh H.W., 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 H.W., Lüdtke O., Nagengast B., Morin A.J.S., 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 H.W., Morin A.J.S., 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 M.M., Roberts B., Gremillion M., von Eye A., Nigg J.T. (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 R.P. (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 A.J.S., Arens A.K., Marsh H.W. (in press-a) A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling.*, DOI: 10.1080/10705511.2014.961800

Morin A.J.S., Maïano C. (2011) Cross-validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM). *Psychology of Sport & Exercise*, **12**(5), 540–554, DOI: 10.1016/j.psychsport.2011.04.003

Morin A.J.S., Marsh H.W. (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 A.J.S., Marsh H.W., Nagengast B. (2013) Exploratory structural equation modeling. In

Hancock G.R., Mueller R.O. (eds) Structural Equation Modeling: A Second Course, pp. 395–436, Greewich, CT, IAP.

Morin A.J.S., 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 A.J.S., 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 A.L., 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 B.O. (2004) Mplus Technical Appendices, Los Angeles, CA, Muthén and Muthén. http://www.statmodel.com/techappen.shtml

Muthén L.K., Muthén B.O. (2012) Mplus User's Guide. Seventh Edition, Los Angeles, CA, Muthén & Muthén.

Nunnally J.C., Bernstein I.H. (1994) Psychometric Theory, New York, McGraw-Hill.

Reininghaus U., Priebe S., Bentall R.P. (2013) Testing the psychopathology of psychosis: evidence for a general psychosis dimension. *Schizophrenia Bulletin*, **39**(4), 884–895, DOI: 10.1093/schbul/sbr182

Reise S.P. (2012) Invited paper: the rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, **47**(5), 667–696, DOI: 10.1080/00273171.2012.715555

Reise S.P., Moore T.M., 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 S.P., Morizot J., Hays R.D. (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 T.A., Sass D.A. (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 L.J., Grös D.F., 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 C.Y. (2002) Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes, Los Angeles, CA, University of California.

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

Supporting information

Online supplements may be accessed with the online version of this article at the publisher's web site.