What are we measuring with the morningness–eveningness questionnaire? Exploratory factor analysis across four samples from two countries

Sareh Panjeh , Sabine Pompeia , Simon N. Archer , Mario Pedrazzoli , Malcolm von Schantz & Hugo Cogo-Moreira

View supplementary material

Published online: 29 Sep 2020.

Submit your article to this journal

Article views: 713

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

ə OPEN ACCESS | Check for updates

What are we measuring with the morningness–eveningness questionnaire? Exploratory factor analysis across four samples from two countries

Sareh Panjeh [a], Sabine Pompeia [b], Simon N. Archer [c], Mario Pedrazzoli [d], Malcolm von Schantz [c], and Hugo Cogo-Moreira [a]

aDepartamento de Psiquiatria, Universidade Federal de São Paulo, São Paulo, Brazil; bDepartamento de Psicobiologia, Universidade Federal de São Paulo, São Paulo, Brazil; cFaculty of Health and Medical Sciences, University of Surrey, Guildford, UK; dEscola de Arte, Ciências E Humanidades, Universidade de São Paulo, São Paulo, Brazil

ABSTRACT

Individual variability in diurnal preference or chronotype is commonly assessed with self-report scales such as the widely used morningness–eveningness questionnaire (MEQ). We sought to investigate the MEQ's internal consistency by applying exploratory factor analysis (EFA) to determine the number of underlying latent factors in four different adult samples, two each from the United Kingdom and Brazil (total $N = 3,457$). We focused on factors that were apparent in all samples, irrespective of particular sociocultural diversity and geographical characteristics, so as to show a common core reproducible structure across samples. Results showed a three-factor solution with acceptable to good model fit indexes in all studied populations. Twelve of the 19 MEQ items in the three-correlated factor solution loaded onto the same factors across the four samples. This shows that the scale measures three distinguishable, yet correlated constructs: (1) items related to how people feel in the morning, which we termed efficiency of dissipation of sleep pressure (recovery process) (items 1, 3, 4, 5, 7, 9, 13, and 19); (2) items related to how people feel before sleep, which we called sensitivity to buildup of sleep pressure (items 2, 10, and 12); and (3) peak time of cognitive arousal (item 11). Although the third factor was not regarded as consistent since only one item was common among all samples, it might represent subjective amplitude. These results suggested that the latent constructs of the MEQ reflect dissociable homeostatic processes in addition to a less consistent propensity for cognitive arousal at different times of the day. By analyzing answers to MEQ items that compose these latent factors, it may be possible to extract further knowledge of factors that affect morningness–eveningness.

Introduction

The alternations between wakefulness and sleep are negotiated through an interaction between the circadian clock (Process C) and a sleep homeostat which measures the buildup and dissipation of sleep pressure (Process S) (Borbély 1982; Dijk and Lockley 2002). A wide range of molecular and behavioral processes, including hormone levels, core body temperature, and sleep-wake patterns, follow intrinsically generated cycles of approximately 24 h, which are known as circadian rhythms. In humans, circadian typology, which produces continuous variables that can be used to classify people into categories or chronotypes (e.g., morning, intermediate, and evening type) with distinguishable morningness–eveningness (M–E) profiles, is one of the most studied individual difference in circadian rhythmicity. However, the idea that M-E reflects a continuum is today regarded as a more appropriate way of

characterizing individuals than using coarsely categorized cutoff chronotype scores (e.g., Caci et al. 2008; Di Milia et al. 2013).

A number of instruments have been proposed to measure individual differences in sleep-wake cycles. These include (number of citations in Web of Knowledge as per 14 June 2020 appended in parenthesis as a measure of their use to date) the morningness–eveningness questionnaire (MEQ: Horne and Östberg 1976, N = 2445 citations), the Composite Scale of Morningness (CSM: Smith et al. 1989, N = 620 citations), which contains 9 items derived from the MEQ and four items from the Diurnal Type Scale (DTS: Torsvall and Åkerstedt 1980, N = 277 citations), the reduced 5-item version of the MEQ (rMEQ: Adan and Almirall 1991, N = 246 citations), the Early/late Preference Scale (PS: Smith et al. 2002, N = 127 citations), the Munich ChronoType Questionnaire (MCTQ: Roenneberg et al. 2003, N = 885 citations), the

Chronotype Questionnaire (ChQ: Ogińska 2011, N = 39 citations), later modified to build the Caen Chronotype Questionnaire (CCQ: see Dosseville et al. 2013, 29 citations), with further modifications by Ogińska et al. (2017), the Circadian Energy Scale (CIRENS: Ottoni et al. 2011, N = 29 citations), and the Morningness–Eveningness Stability Scale-Improved (MESSi: Randler et al. 2016, N = 37 citations), composed of four items derived from the CSM, 9 items from the CCQ (Dosseville et al. 2013), and one item from the CIRENS (Ottoni et al. 2011). With the exception of the MCTQ, which measures the extent to which rhythmic biobehavioural events correspond to environmental ones (phase of entrainment, see Levandovsky et al. 2013; Roenneberg 2015), the other questionnaires evaluate subjective diurnal rhythm, subjective phase, and M–E orientation, correlating with sleep/wake habits (see Oginska 2011), mostly using ordinal, Likert scale-type responses for questions or statements. Although most of these scales are similar in the latter respect, they use different phrasing, number of items and of alternative responses, all of which yield different and non-directly comparable psychometric properties.

The MEQ is the oldest, most cited, and one of the most predominantly used measures in chronobiology and sleep research despite its self-reported and thus subjective nature (see Adan et al. 2012 for a comprehensive review). This is possibly so because scores in the MEQ have been shown to be valid by predicting objective and subjective measures, such as intrinsic circadian period (Duffy et al. 2001), alternations in temperature (Bailey and Heitkemper 2001) and cognition throughout the day (Yoon et al. 1998), circadian secretion of melatonin (Kantermann et al. 2015) and cortisol (Bailey and Heitkemper 2001), patterns of reported sleep and wake times (Di Milia et al. 2013), and capacity to adapt to night shifts and presence of sleep disorders (Sack et al. 2007). This questionnaire enquires about preferred rather than actual timings, and scores vary between individuals due to sociocultural factors (Biswas et al. 2014; Natale et al. 2009), age (Adan and Almirall 1990; Duffy et al. 2001), sex (Adan and Natale 2002), heritability (Hida et al. 2014; von Schantz et al. 2015), and geographical characteristics, such as latitude (e.g., Miguel et al. 2017) and longitude (e.g., Shawa and Roden 2016).

The MEQ was designed to measure M-E to determine suitability for shift work and was initially validated using changes in sleep temperature that are known to follow circadian rhythms (Horne and Östberg 1976). Later studies, however, showed that the MEQ also harbors dimensions related to sleep homeostasis (Mongrain et al. 2006; Taillard et al. 2003; Viola et al. 2007).

A slower build-up of sleep pressure can result in a later preferred bedtime, while a faster dissipation of sleep pressure may favor earlier wake-up time and better general disposition in the morning (Mongrain et al. 2006). This, in part, explains why adolescents, who display eveningness preference in the MEQ and attend schools with early morning start times, achieve insufficient sleep time on school-day nights (Arrona-Palacios et al. 2015). This reflects the earlier developmental reduction of sensitivity to buildup of sleep pressure, leading to a delayed sleep propensity at this age. On the other hand, later maturation of the efficiency of dissipation of sleep pressure, results in the considerable need of adolescents for sleep and, therefore, later wake-up times than an early school start allows (see Crowley et al. 2018). The different ontogenetic time course of sensitivity to buildup and efficiency of dissipation of sleep pressure indicates the physiological separability of these two homeostatic phenomena (Crowley et al. 2018).

Furthermore, subjective amplitude, also termed distinctiveness, is another aspect of sleep-wake cycles, which is explicitly measured in only some of the scales that assess circadian rhythms (i.e., ChQ, CCQ, CIRENS, and MESSi). It reflects the degree of awareness of one's own states of hyper- and hypo-activation and of the ability to modulate psychophysiological state, such as levels of alertness/energy throughout the day or night, representing the range of diurnal variations in mood and activation (Dosseville et al. 2013; Oginska et al. 2017). Because MEQ scores correlate with the subjective amplitude subscale of the ChQ (Dosseville et al. 2013), it is possible that, to some degree, the MEQ also measures circadian amplitude and the homeostatic drive, as discussed above, beyond its ability to assess individual differences in M-E.

Hence, a deeper look into the psychometric properties of the MEQ is necessary to understand more about what the scale actually measures. One of the attributes of the MEQ is its reliability, which has been confirmed by calculating its internal consistency using Cronbach's alpha in many samples worldwide with different characteristics (e.g., Li et al. 2011). However, this index of internal consistency is based on rigid, somewhat unrealistic assumptions, that limit its use under many conditions (Mcneish 2017; Raykov 1997, 1998, 2001), including its applicability to the MEQ. One of these assumptions is that all items of the scale must discriminate the intended measured behavior equally; in other words, the magnitude of the correlation of each item with the underlying M-E indicator should be the same. There is no evidence that this is so in the case of the MEQ. Another assumption is that unidimensionality is

presumed, meaning that all items of the scale should reflect a single construct, which does not seem to be the case. Di Milia et al. (2013), for instance, have pointed out that the homogeneity range among items of the MEQ is low, suggesting the scale is not unidimensional.

Assuming the unidimensionality of the MEQ would mean that M-E is guided by a single process. However, to understand individual differences in this respect, it must be considered that these cycles reflect the interplay between Process C, which depends on time of day (circadian time), and Process S, determined by the duration of wakefulness (Borbély 1982; Dijk and Lockley 2002) and efficiency in dissipating this need during sleep (Rusterholz et al. 2016), as explained above. Additionally, some individuals display what has been termed a bimodal response, characterized as simultaneously signs of extreme morning and evening types (Martynhak et al. 2010; Tempaku et al. 2017). These findings point to the existence of different facets of M-E that may interfere with how people answer different items of the MEQ, which can be masked when total scores of the questionnaire are used, and/or its unidimensionality is assumed.

How these homeostatic and circadian factors affect the way people self-rate their M-E with the MEQ has seldom been investigated. Some studies have proposed to do so by analyzing the MEQ via principal component analysis (PCA), one of the most frequently used model-based component extraction methods (e.g., Jankowski 2013). However, results have been mixed. For instance, Adan and Natale (2002) used the principal component extraction method to determine the MEQ dimensions in an Italian sample and reported three dimensions: Time of greatest efficiency (items 6, 11, 15, 17, 18), sleep time/sleep phase (1, 2, 10, 12, 14, and 16), and awakening time/sleep inertia (3, 4, 5, 7, 8, 9, and 13). By contrast, Li et al. (2011) found two dimensions in a Chinese sample: "sleep phase" (items 1, 2, 3, 4, 5, 6, 7, 8, 13, and 19) and "time of greatest efficiency" (items 10, 11, 12, 14, 15, 16, 17, and 18). Hätönen et al. (2008) used another statistical approach, a factor analysis with least squares and maximum likelihood (ML) estimators. Factor analysis yielded four factors in these Finnish participants: General preferences (items 9, 11, 15, 16, 17, and 19), morning activities (items 1, 4, 5, 7, 8, and 13), evening activities (item 15); and times for physical work (item 12). However, ML estimators are not ideal when answers are not on a continuous scale, such as in the MEQ, and, therefore, alternative estimators that are appropriate for analyzing ordinal scales should have been used, such as weighted least square with mean and variance adjusted (WLSMV) ones (Beauducel and Herzberg 2006).

Clearly, no pattern of aggregation of items into components is apparent from the above-mentioned studies. This may be due to two circumstances. The first is the use of diverse populations, which might present different characteristics that affect how persons answer the MEQ, as discussed above. The second is the inadequacy of PCA for proposing the structural validity of the MEQ, that is, to determine possible different underlying processes affecting the MEQ scores. Indeed, PCA is a variable reduction technique (Brown 2015), which takes the scores of large sets of measured variables (e.g., questions or items of a questionnaire) and reduces them to scores of a smaller set of composite variables called components, which preserve information from the original variables as much as possible and that are not linearly correlated with each other. The first component explains the highest variance in responses to a scale, that is, it accounts for as much variability as possible from the total answers. Each succeeding component then explains progressively smaller amounts of variance, whereas one should decide the limit between the number of meaningful components to be retained, determining if a set of items of a given questionnaire/scale can be represented more parsimoniously by a set of derived components, where the number of components should be fewer than the number of items. Another important issue is that although PCA provides the variance accounted for by each component, these are usually inflated values because errors are not accounted for (Schmitt 2011). Measurement errors can be random, such as arbitrarily choosing a specific alternative answer, misreading a question, or having responses biased by one's present state of mind. In these cases, the errors are unlikely to be the same among different respondents. Nonrandom errors can also occur and be similar among respondents, and these have to do with ambiguities in the wording of a given question, biased answers toward the end of a questionnaire due to respondents' tiredness, etc. As measurement errors increase, reliability decreases, therefore, they have to be taken into account when determining the factor structure and consistency of a scale (Edwards and Bagozzi 2000).

According to Bandalos (2018), factor analysis provides evidence based on the internal consistency (formerly called construct validity). As suggested by Di Milia et al. (2013), a better alternative to describe the structure of the MEQ is to use exploratory factor analysis (EFA), a technique of structural equation modeling (SEM) which explores the structure of the data and returns the number of latent factors that emerge. In fact, latent factors are different from principal components. They are mathematically inferred from scores in items of a questionnaire that share variance and reflect

underlying, non-observable constructs (Brown 2015). If only one factor emerges, for instance, this means that the questionnaire is unidimensional. Two factors would indicate that there are two underlying sources that affect how people answer the questions within a scale. EFA provides factor loadings of each question or item in each latent factor and the pattern of intercorrelations among factors. Importantly, it accounts for measurement errors, unlike PCA (Schmitt 2011). Depending on the way the analysis is set up, factors can be allowed to correlate (oblique rotation) or not (orthogonal rotation) with each factor. It is expected that factors do correlate within a scale such as the MEQ, because all questions track different underlying facets of sleep-wake cycles. Considering that these facets are not correlated would be an unrealistic assumption. In this respect, EFA is also more adequate than PCA, especially when items are in ordered-categorical format (i.e., Likert scale) as in the MEQ.

In sum, although PCA and EFA apparently may create similar results, they are conceptually different. Indeed, only EFA statistically illuminates the most adequate number of latent constructs and the factor structure that underlies a set of variables while considering underlying measurement errors (Brown 2015), the ordinal, and the intercorrelated nature of responses, as is the case of the MEQ. Whenever the purpose is to identify the unknown latent constructs, such as in the case of the MEQ, applying the EFA is the more sensible approach (Fabrigar et al. 1999), since it provides the true evidence of internal consistency.

The only study that has analyzed the factor structure of the MEQ in this way was published by Caci et al. (2008). These authors found a four-factor solution with an exploratory PCA followed by EFA using data of French participants. They named the factors peak time (question/items 1, 9, 11, 15, 17 and 18), morning affect (items 4, 5 and 7), retiring (items 2, 8, 10, 12, 16 and 19), and rising (items 3 and 13). Their choice of software (i.e., STATA), which uses maximum likelihood (ML) as an estimator, might have produced bias in the parameter estimation (factor loadings and thresholds) due to the non-continuous answer formats (i.e., Likert scale), as in the MEQ, similarly to the case of Hätönen et al. (2008). Different authors have argued in favor of estimators from weighted least-squares family in detriment of the ML estimator approach under ordered-categorical items (Beauducel and Herzberg 2006; Kaplan 2012; Muthén and Kaplan 1985). In point of fact, the importance of using factor analysis (i.e., EFA and CFA) has been acknowledged in the analyses of other M–E scales, such as the rMEQ, MESSi and CSM, as will be described below. Therefore, in order to gain insights about the

possible factor structure of the MEQ, we considered the factor structure of other similar questionnaires that were determined with adequate analyses (EFA using WLSMV estimator, scree test, and oblique rotation, CFA or bifactor models).

The psychometric properties of the rMEQ (Adan and Almirall 1991) have been assessed using latent variable approaches (i.e., EFA and CFA), which are robust techniques for understanding the underlying factor structure to a set of items. Urbán et al. (2011), for instance, confirmed a one factor solution proposed by Adan and Almirall (1991). Since the rMEQ has only five items of the MEQ, its unidimensional factor structure is not applicable to that of the full MEQ version.

Regarding the MESSi, some studies have also used latent variable approaches (EFA and CFA), and confirmed a three-factor structure (e.g., Tomažič and Randler 2018). Although four items of the MESSi are derived from the CSM (CSM items 3, 4, 12, and 13), only the first two are derived from the MEQ (MEQ items 4, 5). Hence, MESSi's factor structure is not directly applicable to that of the MEQ.

The third scale analyzed with factor analytic approaches was the CSM, which, at first glance, should have more bearings on the factor structure of the MEQ, because nine of its 13 items (items 1–9) derived from the MEQ (corresponding to items 1, 2, 4, 5, 7, 9, 10, 11, and 19) and also due to the high correlation of total scores (i.e., > 0.90) between the MEQ and CSM (Caci et al. 2008; Smith et al. 1989). However, it must be stressed that the MEQ items within the CSM will not behave psychometrically in the same way, because different patterns of intercorrelations between items might emerge, and consequently, the number of factors and how the items are correlated with them might also change. Overall, the great majority of studies on the CSM show two (Bhatia et al. 2013; Díaz-Morales and Sanchez-Lopez 2004; Kato et al. 2019; Smith et al. 2002) or three (Adan et al. 2005; Caci et al. 2005, 2008; Randler and Díaz-Moralez 2007) interrelated factors that varied in terms of number of items, the items, themselves, in each factor, and also regarding interpretations of what they represent (see Supplementary file for a description of the findings of these papers).

However, these data serve as an indication that M-E does not fit as a single dimension, as indicated by findings of Di Milia et al. (2013). Consequently, we hypothesized that the MEQ would be multidimensional, although we did not predict the number of factors or the pattern of interrelations between the items and factors. Additionally, because various prior studies using the CSM seem to have shown a factor with a large proportion of items that pertains to how people feel in the

morning (Adan et al. 2005; Caci et al. 2005; Kato et al. 2019; Randler and Diaz-Moralez 2007; Smith et al. 1989), we hypothesized that we would also find a factor with this characteristic. Prior CSM and MESSi structure investigations have provided no specific interpretation about what the different factors could represent in terms of homeostatic, subjective amplitude, and/or circadian underlying biological underpinnings that are known to be dissociable using other methodological approaches. This is important because using an adequate statistical approach is not enough to understand the factor structure of a scale without incorporating substantive theory of what the factors indicate. An example would be factors that correspond to separable homeostatic processes (Crowley et al. 2018; Dijk and Lockley 2002; Mongrain et al. 2006; Taillard et al. 2003; Viola et al. 2007).

We, therefore, sought to determine the internal consistency of the MEQ by describing its factor structure using EFA with WLSMV in four samples drawn from two countries, Brazil and the United Kingdom. Although the diversity of these samples in terms of geographical location, sociocultural, and genetic profiles can affect the MEQ answers as discussed above, the rationale behind the use of data from different contexts was to determine a common core structure underlying possible emerging latent factors that reflect sleep patterns that are present irrespective of this diversity, which can nevertheless impacts how people behave (e.g., Rad et al. 2018). This is a first step toward dealing with comparability of models across different samples, so that future studies can use robust types of analyses across samples, such as invariance techniques, which allow direct comparisons of the effects of culture, sex, age, and so forth. In other words, obtaining similar models in samples drawn from different locations is important to produce findings that are not sample-specific and to provide clues about how they vary according to economic and sociocultural factors and context. Having two independent samples each from Brazil and the UK also enabled the determination of variations of responses within the same nation.

Materials and methods

Participants

We used data from four different studies with a total of 3,457 adult participants of both sexes (1,636 men), ranging in age from 18 to 113 y. Study one was conducted in São Paulo, a Brazilian megalopolis (N = 1,177; 425 men; ages ranging from 18 to 80 y; mean = 30.19; [SD = 10.24] y; latitude −23.6; longitude: 46.6). Study two also involved another Brazilian cohort from the small countryside town with less than 20,000 inhabitants, Baependi (N = 1343; 782 men; ages ranging from 18 to 113 y; mean = 45.07 [SD = 17.02] y; latitude: −22.0; longitude: 44.9). Study three and four were collected from the visitors to the Science Museum in London, UK in 2001 and 2004, (2001: N = 484; 217 men; ages ranging from 18 to 78 y; mean = 34.77 [SD = 12.75] y; 2004: N = 453; 212 men; ages ranging from 18 to 81 y; mean = 37.83 [SD = 14.71] y; latitude: 51.5; longitude: 0.2).

Measure of morningness–eveningness

For the UK samples, the original English version of the morningness–eveningness questionnaire (MEQ) (Horne and Östberg 1976) was used, and for the Brazilian samples, the validated translation into Brazilian Portuguese (Benedito-Silva et al. 1990). This questionnaire includes 19 items evaluated on 4 or 5-point ordinal scales generating a score associated with people's morningness (higher scores) or eveningness (lower scores) after reversal of scores of some items. However, we analyzed answers to each item without reversing scores and considered ordinal scores following the instructions from the original questionnaire for the five items (questions 1, 2, 10, 17, and 18) that are rated on a continuous scale.

Procedure

All studies from which the samples were obtained were approved by local Ethics Committees, and all participants provided informed consent, conforming to international ethical standards (Portaluppi et al. 2010). Participants also provided demographic information and filled in the full version of the MEQ. The London samples were collected in person during two separate periods from visitors to the London Sciences Museum (Jones et al. 2007; Robilliard et al. 2002). The São Paulo sample was collected online during two consecutive years as a subset of a larger 13 k sample of the Brazilian population (Miguel et al. 2017). The Baependi sample was collected as part of the Baependi Heart Study through verbal interviews with a scribe (von Schantz et al. 2015). All participants were aged ≥ 18 y. Anonymized databanks from all four studies were aggregated for statistical analyses.

Statistical analysis

Descriptive statistics (proportion and counts at item levels) and polychoric correlation were performed for each sample. Exploratory Factor Analysis (EFA) was

applied separately for each of the four datasets. We employed a weighted least-squares analysis using a diagonal weight matrix with standard errors and mean- and variance-adjusted (WLSMV) estimator. This is the default estimator in Mplus under ordinal items (Muthén and Muthén 1998-2017), and it estimates the magnitude of the loadings more precisely (Beauducel and Herzberg 2006). Oblique rotations were used (GEOMIN), which allow for correlation among latent factors because there was no *a priori* reason to assume that factors would be totally independent. By default, a GEOMIN rotated solution is provided in Mplus (Muthén and Muthén 1998-2017), which is an oblique type of rotation.

Scree test as a first inspection of the number of factors to be extracted was used based on eigenvalues. Scree tests provide a graph in which the factors form the horizontal axis and the eigenvalues create the vertical axis. The graph is used to show the point where lines drawn through the plotted eigenvalues change slope, or the last significant decreasing trend for eigenvalues (Brown 2015). Together with scree tests, standardized solutions were evaluated taking into account the theoretical background underlying the MEQ answers. As the importance of items with small factor loadings, which <0.4, can easily be over-interpreted, we focused on factor loading of each item that were > 0.4, indicating that they had acceptable correlations with their latent factors (Nunnally 1967).

Model fit indices were evaluated as described by Schermelleh-Engel et al. (2003): p-value of χ^2-test, comparative fit index (CFI), Tucker–Lewis index (TLI), root-mean-square error of approximation (RMSEA), and standardized root-mean-square residual (SRMR). The CFI and TLI should be ≥ 0.95. RMSEA values ≤ 0.08 indicate an acceptable to good approximate model fit, while for SRMR, this should be < 0.10. The p-value of the corresponding test of approximate fit (chi-square) should be > 0.05 (i.e., non-statistically significant p-value is considered). Finally, we compared the factor solutions in all samples and the items that loaded in each factor that were common to all studied populations.

Results

Descriptive statistics

Table 1S (supplementary file) shows the proportions and counts according to category of answers of each item per sample. Table 2S (supplementary file) shows the polychoric correlation between items.

Exploratory factor analysis

EFA yielded similar intercorrelated three-factor solutions based on scree test with good fit indices in all samples (Figure 1). The GEOMIN rotated factor loadings, which serve to determine item reliability, and the correlations among factors are shown in Table 1. Table 2 shows the model fit indices across the four tested samples regarding EFA results for one- to four-factor solutions. It is important to consider that, statistically, the more factors that are extracted in the EFA, the better the fit indexes are expected to be, as shown in Table 2 for the four-factor solution across all samples. However, choice
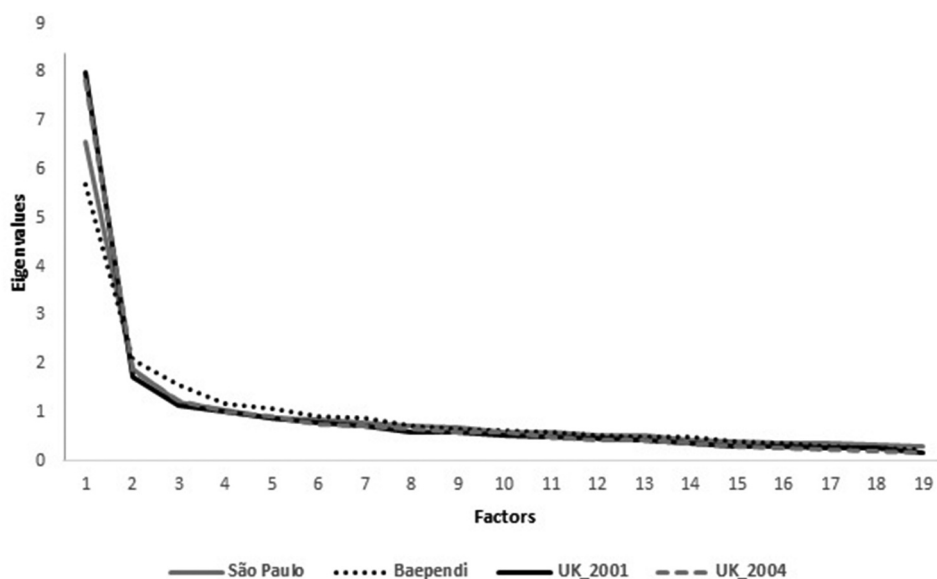
Figure 1. Scree plot of Eigenvalues of the four tested samples. After three factors, there was very little reduction in the Eigenvalues in all four samples, indicating three to be an adequate number of factors.

Table 1. Factor loadings of the three-correlated factor model solution of each Morningness–Eveningness Questionnaire (MEQ) item across the four samples found in the exploratory factor analysis, as well as intercorrelations among factors. Factor loadings higher than 0.4 are marked in bold. Underlined values are replicated with convergent values across all samples.

Items	São Paulo (Brazil) (N = 1,177)			Baependi (Brazil) (N = 1343)			UK2001 (N = 484)			UK2004 (N = 453)		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
1. Preferred waking hour	**0.404**	0.265	0.173	**0.549**	0.177	0.025	**0.682**	0.143	0.057	**0.664**	0.187	−0.018
2. Preferred go-to-bed hour	0.196	**0.682**	0.009	0.129	**0.816**	−0.320	**0.501**	**0.548**	0.005	**0.484**	**0.474**	0.029
3. Dependence on alarm clock	**0.852**	0.045	−0.339	**0.677**	0.024	−0.216	**0.807**	0.005	−0.204	**0.630**	0.025	−0.281
4. Ease in getting up	**0.939**	−0.022	−0.231	**0.928**	−0.049	−0.119	**0.873**	−0.195	−0.009	**0.905**	−0.252	−0.088
5. Alertness after waking up	**0.794**	−0.182	−0.014	**0.732**	−0.045	0.015	**0.833**	−0.343	0.000	**0.878**	−0.406	−0.003
6. Appetite after waking up	0.332	0.036	−0.003	−0.071	0.009	0.045	**0.535**	−0.039	−0.217	**0.409**	−0.032	−0.036
7. Tiredness after waking up	**0.798**	−0.226	0.032	**0.661**	−0.151	0.000	**0.881**	−0.405	0.009	**0.924**	−0.419	0.007
8. Difference preferred/usual time to go to bed	0.231	0.389	−0.103	**0.419**	0.187	−0.073	**0.414**	0.221	−0.078	**0.468**	0.159	−0.151
9. Desire for physical exercise at 07:00–08:00 h	**0.412**	−0.002	0.303	**0.448**	−0.006	0.290	**0.442**	−0.306	0.336	**0.734**	−0.215	0.094
10. Time in the evening when one needs to sleep	0.006	**0.779**	−0.001	−0.009	**0.928**	−0.401	0.204	**0.561**	0.117	0.303	**0.629**	−0.032
11. Time of mental performance	0.000	0.167	**0.768**	0.002	0.117	**0.751**	0.023	−0.028	**0.847**	**0.565**	0.014	**0.486**
12. Level of tiredness at 23:00 h	−0.101	**0.551**	−0.028	−0.341	**0.606**	−0.002	−0.003	**0.461**	0.378	0.189	**0.552**	0.076
13. Sleep length when going to bed several hours later	**0.639**	0.000	−0.194	**0.566**	0.087	−0.166	**0.596**	0.092	−0.145	**0.614**	−0.006	−0.214
14. Sleep when remaining awake between 04:00 and 06:00 h	0.110	0.249	−0.037	0.313	0.213	0.004	0.348	0.136	0.070	0.302	0.284	0.046
15. Preferred time for hard physical work	0.013	0.204	**0.559**	0.121	0.094	0.661	−0.024	0.009	**0.750**	**0.601**	−0.015	0.312
16. Desire for physical exercise at 22:00–23:00 h	−0.042	**0.491**	0.189	−0.170	**0.566**	0.061	0.110	0.301	0.394	0.001	**0.422**	−0.125
17. Five consecutive preferred working hours	0.154	0.166	0.343	**0.435**	0.186	0.136	0.159	0.014	**0.651**	**0.623**	−0.010	**0.452**
18. Time of day of "feeling best" peak	0.140	0.339	0.350	0.238	0.275	0.159	0.280	−0.004	0.372	**0.406**	0.079	**0.446**
19. Self-assessment of morning/ evening-type	**0.528**	0.335	0.046	**0.504**	0.370	0.065	**0.543**	0.147	0.303	**0.754**	0.109	0161
Correlations												
Factor 1	1.000			1.000			1.000			1.000		
Factor 2	0.450	1.000		0.413	1.000		0.764	1.000		0.378	1.000	
Factor 3	0.662	0.367	1.000	0.337	0.391	1.000	0.328	0.355	1.000	0.238	0.265	1.000

of the number of factors must also consider the Scree plot and their theoretical explainability. Comparing the fit indices and dissociable sleep-wake phenomena, the three-correlated factor solution was selected because of the small reduction in Eigenvalues with more than three factors, because it made theoretical sense (addressed in the Discussion) and showed adequate fit indices across the four samples. All *chi*-square *p* values were < 0.05. Since this metric is sensitive to sample size (Tanaka 1987) other indexes were considered. Acceptable to good fits for three-factors were found for the UK 2001 (RMSEA = 0.048, CFI = 0.984, TLI = 0.976, SRMR = 0.036) and UK 2004 samples (RMSEA = 0.060, CFI = 0.977, TLI = 0.966, SRMR = 0.041). In the São Paulo sample, the same was found for three-factors (RMSEA = 0.047, CFI = 0.977, TLI = 0.967, SRMR = 0.033). In the Baependi sample, TLI and CFI for three-factors were not acceptable, but RMSEA was acceptable and SRMR was good (RMSEA = 0.064, CFI = 0.941, TLI = 0.914, SRMR = 0.049). It is important to note that it is not

a requirement that all indexes exhibit *good fit* in order to accept the model solution. Other elements such as interpretability of the factor solution and parsimony of the solution must be considered. For Baependi, we, therefore, also consider the three-correlated factor solution for our discussion and interpretability.

We named the three-correlated factors efficiency of dissipation of sleep pressure (Factor 1), sensitivity to buildup of sleep pressure (Factor 2), and time of greatest cognitive arousal (Factor 3).

The items in bold font in Table 1 are those with factor loadings > 0.4. This is a minimal factor loading considered as meaningful as commonly used in the literature (Nunnally 1967), because it represents at least 16% of the shared variance between the latent factor onto which the item is loaded. There were 12 items with loadings > 0.4 (underlined in Table 1) that were common among all samples in the same factors. Correlations between factors (Table 1) were < 0.8, indicating that the three factors, despite being associated, also correspond to separable constructs, showing good divergent validity.

Table 2. Model fit indices of the one to four factor solutions applied to data of the Morningness–Eveningness Questionnaire (MEQ) using Exploratory Factor Analyses underweighted least square using a diagonal weight matrix with standard errors and mean- and variance-adjusted (WLSMV) estimator across the four samples. df = degrees of freedom; Acceptable factor loadings are: RMSEA = Root Mean Square Error of Approximation ≤.08; SRMR = Standard Root Mean Square Residual ≤ 0.10; CFI = Comparative Fit Index ≥ 0.95; TLI = Tucker–Lewis Index ≥ 0.95.

Location	Factor	Chi-square	df	p-value	RMSEA	SRMR	CFI\TLI
São Paulo	1	2236.080	152	< 0.001	0.108	0.084	0.846\0.827
	2	875.804	134	< 0.001	0.069	0.047	0.945\0.930
	3	427.215	117	< 0.001	0.047	0.033	0.977\0.967
	4	291.141	101	< 0.001	0.040	0.026	0.986\0.976
Baependi	1	3131.842	152	< 0.001	0.121	0.111	0.724\0.689
	2	1491.685	134	< 0.001	0.087	0.073	0.874\0.839
	3	749.980	117	< 0.001	0.063	0.049	0.941\0.914
	4	505.983	101	< 0.001	0.055	0.039	0.962\0.936
UK 2001	1	942.982	152	< 0.001	0.104	0.082	0.901\0.888
	2	402.398	134	< 0.001	0.064	0.049	0.966\0.957
	3	245.376	117	< 0.001	0.048	0.036	0.984\0.976
	4	163.486	101	< 0.001	0.036	0.027	0.992\0.987
UK 2004	1	1072.075	152	< 0.001	0.116	0.091	0.887\0.873
	2	504.606	134	< 0.001	0.078	0.055	0.954\0.942
	3	307.001	117	< 0.001	0.060	0.041	0.977\0.966
	4	188.289	101	< 0.001	0.044	0.030	0.989\0.982

Discussion

This paper investigated the psychometric properties of the MEQ by applying exploratory structural equation modeling (EFA). If the MEQ had simply measured a unidimensional M-E construct, all the items should have loaded on a single factor, good fit indices should have appeared, and all the factor loading should have exhibited very strong values (i.e., close to 1). The analysis revealed, however, that the best solution had three-correlated factors exhibiting the similar structure in terms of factor loadings and their correlation underlying the factors and consequently, it was possible to identify similarities in the factor nature and their core items reproducible across the four samples from the two countries. Our study was not designed to provide evidence of what these separable factors represent; thus, we can only hypothesize about their possible biological underpinnings. However, two of the latent factors underlying the responses to the MEQ items apparently reflect separable homeostatic processes, and the third dimension may relate to subjective amplitude, as will be discussed.

In all samples, the majority of items (1, 3, 4, 5, 7, 9, 13, and 19) loaded on Factor 1, which we named "efficiency of dissipation of sleep pressure", as these items describe how respondents felt when they wake up in the morning (e.g., "Assuming adequate environmental condition, how easy do you find getting up in the morning?", which was the item that had the highest loading on this factor). Faster dissipation of sleep pressure may favor earlier wake-up time and better disposition in the morning, the opposite being true when dissipation is not as efficient (Mongrain et al. 2006). This corroborates prior work with the CSM in which the most numerous

factor is usually found to be related with questions/statements about how people feel at the beginning of the day (e.g., Adan et al. 2005; Caci et al. 2005; Kato et al. 2019; Randler and Díaz-Morales 2007; Smith et al. 1989). Interestingly, item 19, which requires responders to self-rate themselves as evening, morning, or intermediate subtypes, loaded on Factor 1, suggesting that people tended to determine their own chronotypes depending on their efficiency in dissipation of sleep pressure. Turco et al. (2015) assessed the reliability of item 19 of the MEQ by comparing the results with the full questionnaire, with the time of subjective sleepiness during a waking day and also with real-life sleep timing variables. They found that healthy adults could describe their diurnal preference based on this item. Arrona-Palacios and Diaz-Morales (2016) also used this item in Mexican and Spanish adolescents and corroborated these findings. This has important consequences. For instance, genome-wide association studies have determined genetic influences on chronotype by classifying participants in morning or evening chronotypes (Jones et al. 2019). These studies did not use a full questionnaire, but were based on a single question, essentially corresponding to question 19 of the MEQ, which loaded onto Factor 1. However, our results suggest that this item may relate more to dissipation of sleep pressure (Process S) than circadian parameters (Process C). Laboratory studies have shown this to be the case for the variable number tandem repeat (VNTR) polymorphism in the *PER3* gene, which had been reproducibly associated with MEQ score (Viola et al. 2007).

Results of some studies using the CSM also confirm our findings. As noted by Smith et al. (1989), " … the

first (and, therefore, most important factors) are identified by morning items, and the evening items are clustered only in the less significant factors. This result may reflect that the morning items are more diagnostic than evening items in predicting adjustment to schedule changes". Likewise, Adan et al. (2005) showed most items of the CSM (CSM items 3–6 and 10–12, of which the first 4 correspond to MEQ items 4, 5, 7, 9) loaded on a "morning factor". Similar results were reported by Caci et al. (2005), Kato et al. (2019), and Randler and Diaz-Moralez (2007). Moreover, there is similarity between the content of morning-related factors in different studies of the CSM and MEQ items that loaded on factor 1 in our results (items 1, 3, 4, 5, 7, 9, 13, and 19). Moreover, item 9 of the CSM, which corresponds to item 19 of the MEQ, loaded on morning activity in the factor structure of the CSM proposed by Smith et al. (1989) and Morningness/time of day preference factor in the factor structure of the CSM proposed by Kato et al. (2019). What no publication has done to date is contextualize what a "morning factor" means in terms of M-E profiles. We suggest this factor reflects the responders' ability to dissipate sleep pressure, a homeostatic process that is separable from buildup of sleep pressure, which reflects how sleepy people feel at the end of the day (see Crowley et al. 2018; Dijk and Lockley 2002; Mongrain et al. 2006; Taillard et al. 2003; Viola et al. 2007).

Factor 2 was composed of fewer items (items 2, 10, and 12) in common in all samples, items that enquire about how responders feel in the evening, which we named "sensitivity to build-up of sleep pressure", such that a slower buildup of sleep pressure would result in a later preferred bedtime (Mongrain et al. 2006). Among these items the one with the highest factor loading was "At what time in the evening do you feel tired as a result in need of sleep?" (item 10). Item 16, related to how easy the responders find physical exercise in late evening, also loaded on factor 2 in all except the UK 2001 data, showing this to reflect sensitivity to sleep pressure in many, but not all, samples.

Factor 3 was less consistent across samples, as only one item (item 11) of the MEQ loaded on this factor in all studied populations. This item relates to the timing of greatest cognitive arousal ("You wish to be at the peak of your performance for a test which you know is going to be *mentally exhausting* and lasting for two hours … "). An equivalent item related to physical arousal (items 15) also loaded on Factor 3 in three of the four samples (except UK2004). We have no explanation for this smaller loading in this specific sample, but together with the pattern of effects for item 11, our findings suggest that mental/physical arousal tends to be separable from

buildup and dissipation of sleep pressure. This factor may represent subjective amplitude, thought to correspond to the range of variation in people's ability to modulate their levels of alertness, mood, and activation (see Dosseville et al. 2013; Ogińska 2011; Ogińska et al. 2017). Its inconsistency corroborates prior findings that the psychometric properties of subjective amplitude are still unsatisfactory in the scales that specifically assess it (Dosseville et al. 2013; Ogińska 2011; Ogińska et al. 2017), possibly due to two reasons: It might be as the result of the complexity of this phenomenon (Ogińska et al. 2017), or the fact that the MEQ was not designed to measure circadian amplitude. Nonetheless, the subjective amplitude subscale of the ChQ has been found to correlate (weakly) with MEQ scores (Dosseville et al. 2013). It should be considered that amplitude factors have been identified using the MESSi questionnaire in samples from different countries (e.g., Tomažič and Randler 2018), although these studies reported that the internal consistency (Cronbach's alpha) was low for the amplitude factor compared to the other two dimensions (i.e., morning and evening factors) (Tomažič and Randler 2018).

Item 18 also enquires about peak activity ("at what time of day do you think that you reach your *feeling best peak*?"), but it did not consistently load on any factor, not even factor 3, possibly because it was interpreted differently among responders in both countries. After all, "feeling best" is quite an ambiguous term. Additionally, this item is rated on a continuous scale; therefore, inconsistent results may also have occurred because of inadequate cutoff scores used to separate answers into ordinal responses. For instance, responders are awarded three points if the answer is that they feel best from 10:00 until 17:00 h, and two points if between 17:00 and 21:00 h. This may unrealistically separate peak arousal times and result in inconsistent scores, so these cutoff scores should be reassessed. One way of doing so would be to run EFA or confirmatory factor analysis declaring those items as a continuous variable, which would still work under WLSMV estimator. Moreover, the use of clock times in the MEQ has been criticized by some authors (see Di Milia and Randler 2013; Smith et al. 2002) and led to development of scales without clock time items, such as the PS (Smith et al. 2002) and the Morning Affect Scale (Di Milia and Randler 2013). Similarly, item 17, which also relates to arousal during the day ("suppose that you can choose your own work hours … "), had quite a different pattern of loadings among samples (Factor 1 and 3 in UK2004, Factor 3 in the UK2001 sample, Factor 1 in Baependi, and no factor in the São Paulo sample). This may have arisen because of the differences in the predominance of physical and/

or mental work in each sample and/or because its answers are on a continuous scale divided into ordinal scores that may be inappropriate. Overall, our results corroborate the inconsistency in measures of subjective amplitude.

Intercorrelations among factors per sample were low ($r < 0.450$), except for the higher relation of Factors 1 and 2 in the UK2001 sample ($r = 0.764$). Together with what was discussed above, this suggests that the MEQ predominantly reflects at least two clearly separable, but interrelated, homeostatic aspects of sleep-wake cycles, dissipation and buildup of sleep pressure, respectively. This separability explains why, for instance, MEQ scores can indicate individuals of high efficiency of dissipation of sleep pressure (predominant in morning types), low sensitivity to buildup of sleep pressure (evening types), and also ones regarded as displaying both traits (bimodal types: Martynhak et al. 2010; Tempaku et al. 2017), indicating that these processes are indeed separable, as proposed by Rusterholz et al. (2016). There is also developmental evidence for this. In a recent review, Crowley et al. (2018) showed that sensitivity to buildup of sleep pressure markedly decreases in the early years of adolescence, while dissipation of sleep pressure only matures in early adulthood.

Directly comparing the present factor structure with the only MEQ study that used EFA (Caci et al. 2008) is not advisable, because the analyses may have been biased by the limitations of the software, which is adequate for continuous but not ordinal answer formats (Beauducel and Herzberg 2006; Kaplan 2012; Muthén and Kaplan 1985). Still, there is similarity between the factor structure reported by Caci et al. (2008) and our findings. Items 2, 10, and 12 loaded on the "retiring factor" in their study and on factor 2 (buildup of sleep pressure) in ours. Also items 4, 5, and 7 loaded on their "morning factor", which corresponded to our factor 1 (efficiency of dissipation of sleep pressure).

Regarding the 5-item rMEQ, the items that correspond to items 1, 7, and 19 of the full MEQ in the present study loaded on factor 1, while item 10 loaded on factor 2; item 18 did not load on any of our factors. Despite our study not having been designed to directly compare the MEQ and rMEQ, a descriptive contrast between them shows that these five items were not the most reliable items in our factor structure (i.e., did not have the highest factor loadings) nor were they core items (i.e., did not replicate across the samples). Hence, in terms of presenting a balanced factor structure, our findings indicate the rMEQ does not reflect the same factors tapped by the full MEQ. In other words, rMEQ is partly measuring the efficiency of dissipation of sleep pressure and only includes one item representing

factor 2, which is not enough to statistically specify a second domain (minimally two indicators are needed under a multidimensional solution) (see Bollen 1989; Kenny et al. 1998).

It is noteworthy that some items loaded onto specific factors depending on country. This happened for the item related to appetite in the morning (item 6), and best time to go to bed in the evening (item 2), both of which loaded on Factor 1 only in the UK samples, apart from the common loading on Factor 2 of the latter item in all samples. These differences may be related to sociocultural (e.g., Biswas et al. 2014; Natale et al. 2009), age (e.g., Adan and Almirall 1990), or sex (e.g., Adan and Natale 2002). To find a likely configural factor structure, the same number of factors and where the items are loaded is an initial step for psychometric testing of the "stability" of the constructs across samples (see Van de Schoot et al. 2012). To this end, future studies designed for this specific purpose should examine invariance of the two-factor solution underlying the core cross-sample reproducible items of the MEQ. If the model holds despite these variations in scores, this will indicate the MEQ measures the same underlying constructs irrespective of the sample characteristics, an essential property for any tool to understand human behavior. Given the current results from our EFA study, it seems unlikely that the MEQ will achieve strong levels of invariance, for which intercepts/thresholds and factor loadings are assumed to be equal across samples. Without evidence of invariance testing, direct comparison of latent factors found here across samples will not be trustworthy (Vandenberg and Lance 2000).

Moreover, it should be considered that some items loaded onto more than one factor (cross-loadings) in the UK samples (Table 1). This happened for the item related to best time to go to bed (item 2) in both UK samples. Also, items 11, 17 and 18 showed cross loading in the UK 2004 sample. Sociocultural, age, or sex effects may explain these cross-loading regarding items 2 and 11, because they only appeared in the UK populations. Some other items also loaded very differently across samples. Item 17 loaded onto factor 3 in the UK 2001 data and onto factor 1 in Baependi and the UK 2004 data, with cross-loading in the latter sample, but did not load onto any factor in the São Paulo sample. Item 18 did not load onto any factor, except in the UK 2004 sample, with cross-loading. As Brown (2015) suggests, eliminating items with cross-loading, and future studies could consider these items as "*poorly behaved items*" and remove them.

It should be noted that we did not generate Cronbach's Alpha as a general reliability index of the MEQ, mainly because of two issues among others (see

Mcneish 2017; Raykov 1997, 1998, 2001): (a) an assumption of this index is that the scale be unidimensional, while we found a three-factor solution underlying the MEQ; and (b) we used EFA, under which the traditional reliability coefficients cannot be properly calculated – this is more suitable under the CFA approach, which under our rationale of firstly exploring the factor structure, would have required larger samples, in which half the sample would be used to run EFA and the other half to run CFA. Via EFA, it is possible to determine the reliability of each item, i.e., via factor loadings that show the correlation coefficient between observed variables. We also presented the correlation between the latent factors that also indicate the extent to which the factors are distinguishable from each other. Factor analyses also provide evidence-based internal consistency (Bandalos 2018), and the returned solutions and their congruence across the four samples showed some core similarities that could be seen as a first step for future studies interested in cross-cultural comparison.

Our results clearly show that the MEQ measures a multidimensional construct. However, multidimensionality does not guarantee that subscales can provide meaningful and reliable information about subdomains that are distinguishable from a general underlying construct, which is usually measured with total raw scores.

Even in the presence of multidimensionality, the use of the total scale scores (summing all the items) can be justified (Gustafsson and Åberg-Bengtsson 2010). This means that prior findings based on the total MEQ scores constituted by 19 items might be reevaluated, since a great deal of the MEQ datasets are available, excluding items that were poor in terms of reliability, i.e., small factor loadings, and items that showed many cross-loadings, which may reflect their unspecificificity. After excluding items with low factor loadings, data can also be used to test bifactor models in which separable factors or dimensions are assumed to reflect a global underlying construct following analytic corrections proposed by Eid et al. (2017), using CFA departing from our EFA results considering factors 1 and 2 (factor 3 was not reliable) to provide: a) extra evidence about the adequacy of likely subscores across samples with different characteristics; and b) insightful evidence in understanding consistency, specificity, and reliability for each item. Traditional bifactor analyses have shown anomalous results, such as vanishing specific factors and irregular loading patterns. Eid et al. (2017), (2018), (2020) argue that the application of the traditional bifactor modeling requires a two-level sampling process that is usually not present in empirical studies. In these publications, Eid and colleagues also demonstrate how alternative bifactor models with a G-factor and specific factors can be derived

that are more well-defined for the actual single-level sampling design. Therefore, reanalysis under the authors' bifactor specifications would be necessary in understanding the factor structure of the MEQ under a general underlying factor. Based on our results, the ideal model solution would be that of a two-correlated factor solution constituted by eight items in factor 1 (1, 3, 4, 5, 7, 9, 13, and 19) and three items in factor 2 (2, 10, and 12).

Concerning limitations, it should be pointed out that the data of the samples we used were obtained from studies that varied in terms of experimental design and were not all representative of the populations in their respective countries of origin. Therefore, we cannot be certain the results, i.e., three-correlated factor solution, would be replicated in other populations. Still, the samples were large and from two countries that differ in many respects. We focused on the items that similarly loaded on latent factors found in the exploratory factors analyses to try to disentangle aspects of M-E that are sample-unspecific from those that vary according to economic and sociocultural factors and context (see Rad et al. 2018). Obtaining a factor solution with good fit indexes is the first step to allow investigations of invariance in the MEQ responses worldwide. Invariance analysis ensures that the psychometric properties of a scale are directly comparable across cultures, age, sex, etc., disentangling bias related to item features, i.e., factor loadings and thresholds, that can vary depending (i.e., are not stable) on cross-population characteristics. This type of testing uses CFA procedures that were not carried out here. The reason for this was that although the Brazilian datasets were large, which could have allowed us to split the sample in two and calculate EFA for one half and confirm the factors structure with CFA, for the UK data, the samples were too small considering the number of thresholds to be estimated (alternative answers for all items). If we had done so, we would have faced imprecision of confidence intervals.

In summary, we conclude that the MEQ measures two consistent separable, yet interrelated, latent constructs across four different samples that seem to reflect two homeostatic processes, sensitivity to buildup of sleep pressure and efficiency of dissipation of sleep pressure, rather than circadian parameters. Items that enquire about physical and mental arousal during the day seem to be much more varied according to the population under study, even when comparing samples from the same country, possibly because of difference in culture/context and/or demographic characteristics. This paper is intended to serve as a prototype about how to conduct EFA and the strengths of using

WLSMV as an estimator, and it also proposes ways to improve the psychometric features of the MEQ.

Disclosure of interest

The authors report no conflict of interest.

ORCID

Sareh Panjeh 🔟 http://orcid.org/0000-0001-8335-1524
Sabine Pompeia 🔟 http://orcid.org/0000-0002-5208-8003
Simon N. Archer 🔟 http://orcid.org/0000-0002-9273-4647
Mario Pedrazzoli 🔟 http://orcid.org/0000-0002-5257-591X
Malcolm von Schantz 🔟 http://orcid.org/0000-0002-9911-9436
Hugo Cogo-Moreira 🔟 http://orcid.org/0000-0001-9411-9237

References

Adan A, Almirall H. 1990. Adaptation and standardization of a Spanish version of the morningness-eveningness questionnaire: individual differences. Pers Individ Differ. 11(11):1123–1130. doi:10.1016/0191-8869(90)90023-K

Adan A, Almirall H. 1991. Horne & Östberg morningness-eveningness questionnaire: a reduced scale. Pers Individ Differ. 12(3):241–253. doi:10.1016/0191-8869(91)90110-W

Adan A, Archer SN, Hidalgo MP, Di Milia L, Natale V, Randler C. 2012. Circadian typology: a comprehensive review. Chronobiol Int. 29(9):1153–1175. doi:10.3109/07420528.2012.719971

Adan A, Caci H, Prat G. 2005. Reliability of the Spanish version of the composite scale of morningness. Eur Psychiat. 20(7):503–509. doi:10.1016/j.eurpsy.2005.01.003

Adan A, Natale V. 2002. Gender differences in morningness-eveningness preference. Chronobiol Int. 19(4):709–720. doi:10.1081/CBI-120005390

Arrona-Palacios A, Díaz-Morales JF. 2016. Morningness-eveningness and sleep habits at school: a comparative study between Mexico and Spain. Biol Rhythm Res. 48(2):175–188. doi:10.1080/09291016.2016.1245459

Arrona-Palacios A, García A, Valdez P. 2015. Sleep-wake habits and circadian preference in Mexican secondary school. Sleep Med. 16(10):1259–1264. doi:10.1016/j.sleep.2015.05.026

Bailey SL, Heitkemper MM. 2001. Circadian rhythmicity of cortisol and body temperature: morningness-eveningness effects. Chronobiol Int. 18(2):249–261. doi:10.1081/CBI-100103189

Bandalos DL. 2018. Validity. In: Measurement theory and applications for the social sciences. New York (NY): Guilford Press; p. 254–297.

Beauducel A, Herzberg PY. 2006. On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. Struct Equ Model. 13(2):186–203. doi:10.1207/s15328007sem1302_2

Benedito-Silva AA, Menna-Barreto L, Marques N, Tenreiro S. 1990. A self-assessment questionnaire for the determination of morningness-eveningness types in Brazil. Prog Clin Biol Res. 341B:89–98.

Bhatia T, Agrawal A, Beniwal RP, Thomas P, Monk TH, Nimgaonkar VL, Deshpande SN. 2013. A Hindi version of the Composite Scale of Morningness. Asian J Psychiatr. 6(6):581–584. doi:10.1016/j.ajp.2013.09.001

Biswas A, Adan A, Haldar P, Majumder D, Natale V, Randler C, Tonetti L, Sahu S. 2014. Exploration of trans-cultural properties of the reduced version of the morningness-eveningness questionnaire (rMEQ) using adaptive neuro-fuzzy inference system. Boil Rhythm Res. 45(6):955–968. doi:10.1080/09291016.2014.939442

Bollen KA. 1989. Structural equations with latent variables. New York (NY): John Wiley & Sons.

Borbély AA. 1982. A two process model of sleep regulation. Hum Neurobiol. 1(3):195–204.

Brown TA. 2015. Confirmatory factor analysis for applied research. 2nd ed. New York (NY): Guilford Press.

Caci H, Adan A, Bohle P, Natale V, Pornpitakpan C, Tilley A. 2005. Transcultural properties of the composite scale of morningness: the relevance of the 'morning affect' factor. Chronobiol Int. 22(3):523–540. doi:10.1081/CBI-200062401

Caci H, Deschaux O, Adan A, Natale V. 2008. Comparing three morningness scales: age and gender effects, structure and cut-off criteria. Sleep Med. 10(2):240–245. doi:10.1016/j.sleep.2008.01.007

Crowley SJ, Wolfson AR, Tarokh L, Carskadon MA. 2018. An update on adolescent sleep: new evidence informing the perfect storm model. J Adolesc. 67:55–65. doi:10.1016/j.adolescence.2018.06.001

Di Milia L, Adan A, Natale V, Randler C. 2013. Reviewing the psychometric properties of contemporary circadian typology measures. Chronobiol Int. 30(10):1261–1271. doi:10.3109/07420528.2013.817415

Di Milia L, Randler C. 2013. The stability of the morning affect scale across age and gender. Pers Individ Diff. 54(2):298–301. doi:10.1016/j.paid.2012.08.031

Díaz-Morales JF, Sánchez-López MP. 2004. Composite and preferences scales of morningness: reliability and factor invariance in adult and university samples. Span J Psychol. 7(2):93–100. doi:10.1017/s1138741600004790

Dijk DJ, Lockley SW. 2002. Integration of human sleep-wake regulation and circadian rhythmicity. J A Physiol. 92(2):852–862. doi:10.1152/japplphysiol.00924.2001

Dosseville F, Laborde S, Lericollais R. 2013. Validation of a chronotype questionnaire including an amplitude dimension. Chronobiol Int. 30(5):639–648. doi:10.3109/07420528.2012.763042

Duffy JF, Rimmer DW, Czeisler CA. 2001. Association of intrinsic circadian period with morningness-eveningness, usual wake time, and circadian phase. Behav Neurosci. 115(4):895–899. doi:10.1037//0735-7044.115.4.895

Edwards JR, Bagozzi RP. 2000. On the nature and direction of relationships between constructs and measures. Psychol Methods. 5(2):155–174. doi:10.1037/1082-989x.5.2.155

Eid M. 2020. Multi-faceted constructs in abnormal psychology: implications of the bifactor S-1 model for individual clinical assessment. J Abnorm Child Psychol. 1–6. doi:10.1007/s10802-020-00624-9

Eid M, Geiser C, Koch T, Heene M. 2017. Anomalous results in G-factor models: explanations and alternatives. Psychol Methods. 22(3):541–562. doi:10.1037/met0000083

Eid M, Krumm S, Koch T, Schulze J. 2018. Bifactor models for predicting criteria by general and specific factors: problems of nonidentifiability and alternative solutions. J Intell. 6 (3):42. doi:10.3390%2Fjintelligence6030042

Fabrigar LR, Wegener DT, Maccallum RC, Strahan EJ. 1999. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods. 4(3):272–299. doi:10.1037/1082-989X.4.3.272

Gustafsson J-E, Åberg-Bengtsson L. 2010. Unidimensionality and interpretability of psychological instruments. In: Embretson SE, editor. Measuring psychological constructs: advances in model-based approaches. Washington (DC): American Psychological Association; p. 97–121.

Hätönen T, Forsblom S, Kieseppa T, Lonnqvist J, Partonen T. 2008. Circadian phenotype in patients with the co-morbid alcohol use and bipolar disorders. Alcohol Alcohol. 43 (5):564–568. doi:10.1093/alcalc/agn057

Hida A, Kitamura S, Katayose Y, Kato M, One H, Kadotani H, Uchiyama M, Ebisawa T, Inoue Y, Kamei Y, et al. 2014. Screening of clock gene polymorphisms demonstrates association of a PER3 polymorphism with morningness-eveningness preference and circadian rhythm sleep disorder. Sci Rep. 4(1):1–6. doi:10.1038/srep06309

Horne JA, Östberg O. 1976. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol. 4(2):97–110.

Jankowski KS. 2013. Polish version of the reduced morningness-eveningness questionnaire. Biol Rhythm Res. 4(3):427–433. doi:10.1080/09291016.2012.704791

Jones KHS, Ellis J, von Schantz M, Skene DJ, Dijk DJ, Archer SN. 2007. Age-related change in the association between a polymorphism in the PER3 gene and preferred timing of sleep and waking activities. J Sleep Res. 16 (1):12–16. doi:10.1111/j.1365-2869.2007.00561.x

Jones SE, Lane JM, Wood AR, van Hees VT, Tyrrell J, Beaumont RN, Jeffries AR, Dashti HS, Hillsdon M, Ruth KS, et al. 2019. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. Nat Commun. 10(1):343. doi:10.1038/s41467-018-08259-7

Kantermann T, Sung H, Burgess HJ. 2015. Comparing the morningness-eveningness questionnaire and Munich chronotype questionnaire to the dim light melatonin onset. J Biol Rhythm. 30(5):449–453. doi:10.1177/0748730415597520

Kaplan D. 2012. Structural equation modeling: foundations and extensions. 2nd ed. Thousand Oaks (CA): SAGE Publications.

Kato Y, Urbán R, Saito S, Yoshida K, Kurokawa M, Rigó A. 2019. Psychometric properties of a Japanese version of Composite Scale of Morningness. Heliyon. 5(1):1. doi:10.1016/j.heliyon.2018.e01092

Kenny DA, Kashy DA, Bolger N. 1998. Data analysis in social psychology. In: Gilbert DT, Fiske ST, Lindzey G, editors. The handbook of social psychology. 4th ed. New York (NY): McGraw-Hill; p. 233–265.

Levandovski R, Sasso E, Hidalgo MP. 2013. Chronotype: a review of the advances, limits and applicability of the main instruments used in the literature to assess human phenotype. Trends Psychiatry Psychother. 35(1):3–11. doi:10.1590/S2237-60892013000100002

Li S-X, Li -Q-Q, Wang X-F, Liu L-J, Liu Y, Zhang L-X, Zhang B, Lu L. 2011. Preliminary test for the Chinese version of the morningness-eveningness questionnaire. Sleep Biol Rhythms. 9(1):19–23. doi:10.1111/j.1479-8425.2010.00480.x

Martynhak BJ, Louzada F, Pedrazzoli M, Araujo JF. 2010. Does the chronotype classification need to be updated? Preliminary findings. Chronobiol Int. 27(6):1329–1334. doi:10.3109/07420528.2010.490314

Mcneish D. 2017. Thanks coefficient alpha, we will take it from here. Psychol Methods. 23(3):412–433. doi:10.1037/met0000144

Miguel MAL, Louzada FM, Duarte LL, Areas RP, Alam M, Freire MV, Fontenele-Araujo J, Menna-Barreto L, Pedrazzoli M. 2017. Latitudinal cline of chronotype. Sci Rep. 7(1):5437. doi:10.1038/s41598-017-05797-w

Mongrain V, Carrier J, Dumont M. 2006. Circadian and homeostatic sleep regulation in morningness-eveningness. J Sleep Res. 15(2):162–166. doi:10.1111/j.1365-2869.2006.00532.x

Muthén B, Kaplan D. 1985. A comparison of some methodologies for the factor analysis of non-normal Likert variables. Br J Math Stat Psychol. 38(2):171–189. doi:10.1111/j.2044-8317.1985.tb00832.x

Muthén LK, Muthén BO. 1998-2017. Mplus user's guide. 8th ed. Los Angeles (CA): Muthén & Muthén.

Natale V, Adan A, Fabbri M. 2009. Season of birth, gender, and social-cultural effects on sleep timing preferences in humans. Sleep. 32(3):423–426. doi:10.1093/sleep/32.3.423

Nunnally J. 1967. Psychometric theory. New York (NY): McGraw-Hill.

Ogińska H. 2011. Can you feel the rhythm? A short questionnaire to describe two dimensions of chronotype. Pers Ind Diff. 50(7):1039–1043. doi:10.1016/j.paid.2011.01.020

Ogińska H, Mojsa-Kaja J, Mairesse O. 2017. Chronotype description: in search of a solid subjective amplitude scale. Chronobiol Int. 34(10):1388–1400. doi:10.1080/07420528.2017.1372469

Ottoni GL, Antoniolli E, Lara DR. 2011. The circadian energy scale (CIRENS): two simple questions for a reliable chronotype measurement based on energy. Chronobiol Int. 28 (3):229–237. doi:10.3109/07420528.2011.553696

Portaluppi F, Smolensky MH, Touitou Y. 2010. Ethics and methods for biological rhythm research on animals and human beings. Chronobiol Int. 27(9–10):1911–1929. doi:10.3109/07420528.2010.516381

Rad MS, Martingano AJ, Ginges J. 2018. Toward a psychology of homo sapiens: making psychological science more representative of the human population. Proc Natl Acad Sci. 115 (45):11401–11405. doi:10.1073/pnas.1721165115

Randler C, Díaz-Morales JF. 2007. Morningness in German and Spanish students: a comparative study. Eur J Pers. 21 (4):419–427. doi:10.1002/per.632

Randler C, Díaz-Morales JF, Rahafar A, Vollmer C. 2016. Morningness-eveningness and amplitude-development and validation of an improved composite scale to measure circadian preference and stability (MESSi). Chronobiol Int. 33(7):832–848. doi:10.3109/07420528.2016.1171233

Raykov T. 1997. Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. Multivar Behav Res. 32(4):329–353. doi:10.1207/s15327906mbr3204_2

Raykov T. 1998. Coefficient alpha and composite reliability with interrelated nonhomogeneous items. Appl Psychol Meas. 22(4):375–385. doi:10.1177/014662169802200407

Raykov T. 2001. Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. Brit J Math Stat Psy. 54(2):315–323. doi:10.1348/000711001159582

Robilliard DL, Archer SN, Arendt J, Lockley SW, Hack LM, English J, Leger D, Smits MG, Williams A, Skene DJ, et al. 2002. The 3111 *Clock* gene polymorphism is not associated with sleep and circadian rhythmicity in phenotypically characterized human subjects. J Sleep Res. 11(4):305–312. doi:10.1046/j.1365-2869.2002.00320.x

Roenneberg T. 2015. Having trouble typing? What on earth is chronotype? J Biol Rhythm. 30(6):487–491. doi:10.1177/0748730415603835.

Roenneberg T, Wirz-Justice A, Merrow M. 2003. Life between Clocks: daily temporal patterns of human chronotypes. J Biol Rhythm. 18(1):80–90. doi:10.1177/0748730402239679

Rusterholz T, Tarokh L, Van Dongen HP, Achermann P. 2016. Interindividual differences in the dynamics of the homeostatic process are trait-like and distinct for sleep versus wakefulness. J Sleep Res. 26(2):171–178. doi:10.1111/jsr.12483

Sack RL, Auckley D, Auger RR, Carskadon MA, Wright KP, Vitiello MV, Zhdanova IV. 2007. Circadian rhythm sleep disorders: part I, basic principles, shift work and jet lag disorders. An American academy of sleep medicine review. Sleep. 30(11):1460–1483. doi:10.1111/jsr.12483

Schermelleh-Engel K, Moosbrugger H, Müller H. 2003. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. MPR-Online. 8(2):23–74.

Schmitt TA. 2011. Current methodological considerations in exploratory and confirmatory factor analysis. J Psychoeduc Assess. 29(4):304–321. doi:10.1177/0734282911406653

Shawa N, Roden LC. 2016. Chronotype of South African adults is affected by solar entrainment. Chronobiol Int. 33(3):315–323. doi:10.1177/0734282911406653

Smith CS, Folkard S, Schmieder RA, Parra LF, Spelten E, Almiral H, Sen RN, Sahu S, Perez LM, Tisak J. 2002. Investigation of morning-evening orientation in six countries using the preferences scale. Pers Ind Diff. 32(6):949–968. doi:10.1016/s0191-8869(01)00098-8

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. doi:10.1037/0021-9010.74.5.728

Taillard J, Philip P, Coste O, Sagaspe P, Bioulac B. 2003. The circadian and homeostatic modulation of sleep pressure during wakefulness differs between morning and evening chronotypes. J Sleep Res. 12(4):275–282. doi:10.1046/j.0962-1105.2003.00369.x

Tanaka JS. 1987. 'How Big Is Big Enough?': sample size and goodness of fit in structural equation models with latent variables. Child Dev. 58(1):134. doi:10.2307/1130296

Tempaku PF, Ramirez Arruda J, Mazzotti DR, Gonçalves BSB, Pedrazzoli M, Bittencourt L, Tufik S. 2017. Characterization of bimodal chronotype and its association with sleep: a population-based study. Chronobiol Int. 34(4):504–510. doi:10.1080/07420528.2017.1306707

Tomažič I, Randler C. 2018. Slovenian adaptation of the Morningness-Eveningness-Stability Scales improved (MESSi). Biol Rhythm Res. 51(3):453–459. doi:10.1080/09291016.2018.1535539

Torsvall L, Åkerstedt T. 1980. A diurnal type scale. Construction, consistency and validation in shift work. Scand J Work Environ Health. 6(4):283–290. doi:10.1080/09291016.2018.1535539

Turco M, Corrias M, Chiaromanni F, Bano M, Salamanca M, Caccin L, Merkel C, Amodio P, Romualdi C, De Pittà C, et al. 2015. The self-morningness/eveningness (Self-ME): an extremely concise and totally subjective assessment of diurnal preference. Chronobiol Int. 32(9):1192–1200. doi:10.3109/07420528.2015.1078807

Urbán R, Magyaródi T, Rigó A. 2011. Morningness-eveningness, chronotypes and health-impairing behaviors in adolescents. Chronobiol Int. 28(3):238–247. doi:10.3109/07420528.2010.549599

Van de Schoot R, Lugtig P, Hox J. 2012. A checklist for testing measurement invariance. Eur J Dev Psychol. 9(4):486–492. doi:10.3109/07420528.2010.549599

Vandenberg RJ, Lance CE. 2000. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ Rese Methods. 3(1):4–70. doi:10.3109/07420528.2010.549599

Viola AU, Archer SN, James LMM, Groeger JA, Lo JCY, Skene DJ, von Schantz M, Dijk DJ. 2007. PER3 polymorphism predicts sleep structure and waking performance. Curr Biol. 17(7):613–618. doi:10.3109/07420528.2010.549599

von Schantz M, Taporoski TP, Horimoto ARVR, Duarte NE, Vallada H, Krieger JE, Pedrazzoli M, Negrão AB, Pereira AC. 2015. Distribution and heritability of diurnal preference (chronotype) in a rural Brazilian family-based cohort, the Baependi study. Sci Rep. 5(1):1–6. doi:10.3109/07420528.2010.549599

Yoon C, May CP, Hasher L. 1998. Aging, circadian arousal patterns, and cognition. In: Schwarz N, Park D, Knauper B, Sudman S, editors. cognition, aging and self-reports. 1st ed. New York (NY): Psychology Press; p. 113–136.