



Analysis of the Psychometric Properties of the Glasgow Sleep Effort Scale Through Classical Test Theory, Item Response Theory, and Network Analysis

Sofia Fontoura Dias¹ · Ana Allen Gomes^{2,3} · Colin A. Espie⁴ · Daniel Ruivo Marques^{1,3}

Received: 23 September 2022 / Revised: 11 March 2023 / Accepted: 18 April 2023 / Published online: 28 April 2023
© The Author(s) 2023

Abstract

Purpose The Glasgow Sleep Effort Scale was developed with the aim of measuring sleep effort. The present study evaluates the psychometric properties of the European Portuguese version of the scale through classical test theory (specifically confirmatory factor analysis), item response theory, and network analysis.

Methods It was used an existing database comprising a community sample of 227 Portuguese adults, aged 20–74 ($M = 42.99$; $SD = 12.44$) years old, comprising 49.8% women and 50.2% men.

Results In confirmatory factor analysis, a well-adjusted to data one-factor model was identified, with sleep effort as the single factor. Item response theory analysis indicated an adequate performance of all items and satisfactory coverage of the latent trait, with items 3 ($a = 3.37$), 4 ($a = 3.18$), and 6 ($a = 3.02$) as the most discriminative ones. Considering network analysis, items 4 and 7 presented the strongest edge weight (.48). Item 1 was displayed at the centre of the networks and accounted for the largest number of connections with the remaining items.

Conclusion The results suggest that the Glasgow Sleep Effort Scale is a reliable measure, comprising highly discriminative items with good centrality indices.

Keywords Sleep effort · Psychometrics · Classical test theory · Item response theory · Network analysis

1 Introduction

Insomnia is generally considered the most common sleep disorder, and one of the mental health disorders with higher prevalence rates. As a chronic and clinically significant

disorder, its prevalence is approximately 10%, but nearly 40% of the population reports poor sleep complaints [1]. According to the *International Classifications of Sleep Disorders (ICSD-3)* [2] and the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* [3], insomnia comprises self-reported difficulties in either falling asleep, maintaining sleep, and/or waking-up earlier than the desired time. The mentioned difficulties must be specifically present at least three nights per week, for 3 months, and occur despite adequate circumstances and opportunity to sleep. Dysfunction in daily life must also be addressed in several areas [1].

Within the large set of theories on the aetiology and pathophysiology of insomnia, the hyperarousal construct is often highlighted [4]. According to the psychobiological inhibition model by Espie [5], under normal conditions, sleep occurs without attention, intention, or effort. In the presence of possible stressful life events, there is physiological and psychological arousal. Consequently, the inhibition of sleep-related de-arousal occurs, leading to insomnia symptoms [6].

✉ Daniel Ruivo Marques
drmarques@ua.pt

¹ Department of Education and Psychology, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

² Faculty of Psychology and Educational Sciences, University of Coimbra, Rua Do Colégio Novo, 3000-115 Coimbra, Portugal

³ CINEICC-Centre for Research in Neuropsychology and Cognitive Behavioural Intervention, Faculty of Psychology and Educational Sciences, University of Coimbra, Coimbra, Portugal

⁴ Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

Sleep effort became an important symptom of psychophysiological insomnia, which led to the development of a scale aimed to measure it—the Glasgow Sleep Effort Scale—GSES. The authors, considering pilot versions of the scale, posited a conceptual model of sleep effort with seven main components, where each one was assigned to an item. The GSES is short, of easy application and the initial validation studies demonstrated good indicators of sensitivity, specificity, and discriminant validity [7].

The first validation study of the GSES was conducted by Broomfield and Espie (2005) comprising a sample of insomnia patients and good sleepers (healthy controls). Results from 89 insomnia patients and 102 good sleepers indicated adequate internal consistency values (Cronbach's $\alpha=0.77$) and good ability to differentiate groups. A 2-point cut-off correctly identified 93.3% of the patients and 87.3% of good sleepers. For a 3-point cut-off, 82.2% patients and 92.5% good sleepers were accurately identified by the scale. Recurring to a principal component analysis (PCA), a single component accounted for 62.6% of the variance and each item presented a significant loading on this component with a mean α of 0.79 [7]. In the study by Kohn and Espie [8], encompassing three groups (psychophysiological insomnia patients, insomnia associated with mental disorder patients, and good sleepers), results showed a higher score in the insomnia group associated with other mental disorders, followed by the group with psychophysiological insomnia and the good sleepers.

In 2016, the European Portuguese version of GSES [9] was studied in a young adult sample, with good psychometric properties. Results showed good internal consistency ($\alpha=0.79$). The GSES has been widely used in studies with insomnia patients [10–13], and specifically in insomnia behavioural treatments such as stimulus control-based therapy [14] or sleep restriction therapy [15]. Sleep effort was also associated with sleep disturbance severity in patients with anxiety and depressive disorders [16] and in insomnia and major depressive disorders [17].

Up to now, what is known on the GSES is exclusively based in classical psychometric approaches—Classical Test Theory (CTT). Actually, this is a mainstream practice in self-report measures in sleep medicine and other domains. Although it is not an erroneous practice, evidence for more alternative and contemporary approaches, such as the Item Response Theory (IRT) and Network Analysis (NA), is lacking [18].

When describing psychometric approaches in psychological test development, Factor Analysis and IRT are good examples of what is referred to continuous Latent Variable Models (LVM). Unobservable elements (latent variables) are considered the ultimate cause of one's response to each item of a test. In particular, these are the “factors” in factor analysis and “latent traits” in IRT [19, 20].

Factor analysis relies on correlation patterns [21]. Groups of variables that are highly associated constitute a factor. Factors are described as attributes or dimensions that vary along with different individuals and, consequentially, along a test scale [21, 22]. In psychometrics, factor analysis is important to evaluate construct validity. The dimensional or internal structure can be assessed either in an exploratory or confirmatory manner. In Exploratory Factor Analysis (EFA), it is the data that will guide the composition of the factors [20]. In turn, CFA is related to very specific hypotheses about a test's internal structure, either in the number of factors identified, the relations between factors and the items, and the feasible associations between the factors. Moreover, CFA is often used in later stages of test development, when there is a clearer comprehension about the constructs, the item properties, and the test itself [23]. It is noteworthy to refer that both EFA and PCA share the aim of reducing data sets, despite different theoretical foundations. In PCA, all variance (shared and unique) in the data is analysed, assuming no error. EFA considers only shared variance and error variance is assumed. For this reason, in PCA, the extracted variables are labeled components and not factors [21].

Despite its wide utilization, classical psychometrics present some problematic issues that may be surpassed with other approaches. For instance, CTT techniques often rely on the interpretation of test scores, which can be limited when we can obtain information on the participant performance in each of the items [24]. Therefore, IRT is a set of mathematical and statistical techniques used to estimate important intrinsic parameters of the items [25]. It allows the description of the “relationship between the latent trait, the properties of the items in the scale, and respondents' answers to the individual items” [26] (p. 171). This relationship is best described by a monotonic increasing mathematical equation named *Item Characteristic Curve (ICC)* [27]. In this regard, as the individual's latent trait level (θ) raises, the same pattern is observed in the probability of one's endorsing an item [28]. The main parameters for IRT analysis are item location/difficulty (b), discrimination (a), guessing (c), and trait score. Item location (b), also referred to as item difficulty, is related to the position on theta where the probability of endorsing an item is equal to 50% in the simplest dichotomic case [24, 27, 28]. In polytomic items, several b parameters must be calculated, since they comprise intersection points between two curves. These determine the theta value where the individual achieves the next superior value [24]. The discrimination parameter (a) corresponds to the ICC slope at b . It enables the description of how well an item distinguishes between different participants taking into consideration their levels of latent trait [26, 28]. Finally, the (c) guessing parameter evaluates the probability of responding correctly to an item by chance [24, 28]. To conduct an IRT analysis, one must consider unidimensional models, which are evaluated

with factor analysis [24]. If multiple dimensions come up, researchers must ensure that one-dimension accounts for at least 20% of the variance [28]. Other key assumption is local independency, meaning that each item is statistically independent of responses of the remaining and it is only influenced by theta levels [26]. Further, monotonicity refers to the assumption of increased probability of endorsing an item as the one's latent trait increases [28], a principle present in the ICC [26, 28]. Finally, item invariance accounts item parameters' consistency across distinct populations. When there are differences in item responses between groups, there is differential item functioning (DIF) [28].

More recently, NA emerges as an alternative theoretical approach that is not new in the science domain, but it has just recently gained relevance in psychopathology. Up to this point, this approach is still not extensively used in mental health, despite its particular significance for this context due to the massive number of variables and its interactions [29, 30]. Instead of a latent variable explaining the data (or underlying disorder), it is posited a network of symptoms that are associated—sometimes of causal nature [29]. Psychopathology can be viewed as a system comprising symptoms and signs that is complex, dynamic, and interchangeable [31, 32]. In the basic structure of NA are nodes and edges. Briefly, nodes are the representation of objects or variables that one wants to examine, and edges comprise the relationships between nodes. In psychological networks, nodes are observed psychological variables (e.g., symptoms, behaviours, cognitions, etc.) and edges are unknown statistical associations (e.g. correlations, predictive relationships) [30, 32, 33]. Additionally, a node can also be a representation of a particular scale item, a sub-scale, or even a composite scale and edges can represent relationships such as symptom comorbidity [30]. Furthermore, this approach allows the identification of the most central symptoms that are used to diagnose and plan treatment [32, 34]. It is also noteworthy that, in sleep behaviour, “bridge” symptoms—those which link two sets of networks—are fundamental pieces considering the comorbidity with other psychological and medical disorders [35]. By intervening in specific nodes' subnetworks, and eventually the whole network, those symptoms can be activated or deactivated for a certain duration [34]. Hitherto, there is scarce evidence on the use of this approach in the sleep medicine field [35].

Particularly, in psychometrics, recent evidence has been highlighting the combination of other novel network approaches with psychometric techniques from CTT and IRT for a more accurate evaluation of self-report measures [36]. Moreover, despite a large amount of published literature on reliable and valid measures in clinical and research practice, there has been growing interest in the reduction of the measures' dimension (e.g., the *Insomnia Severity Scale* [ISI-3]) to overcome the burden of patients with time-consuming

clinical evaluations and research protocols [18, 37]. Even though minimal patient overload on the assessment with short measures (such as the ISI-3 when singly used), in full clinical evaluations “each item represents additional burden” [37] (p. 105).

The major aim of the current study is to assess the psychometric properties of the European Portuguese version of the GSES using CTT, IRT, and NA. In an attempt to build upon psychometric research, this study examines if the mentioned psychometric approaches can (or cannot) be complementary.

2 Methods

2.1 Participants

This study is a secondary data analysis using an existing database [38] comprising a non-probabilistic sample of 227 Portuguese adults. The main aim of that study was to assess the psychometric properties of the *Basic Scale on Insomnia complaints and Quality of Sleep* (BaSIQS), the GSES and GCTI in its Portuguese European version. Insomnia complaints and sleep quality, sleep effort, and thoughts when one cannot fall asleep were examined as well. The author tried to recruit the most heterogeneous and representative sample of the Portuguese adult population in the North region of the country. Inclusion criteria were: (a) age of 20 or more years, (b) having or have had a professional occupation, and (c) the voluntary participation in the study.

Participants were aged between 20 and 74 years ($M = 42.99$, $SD = 12.44$). Women represented 50.2% of the sample ($n = 114$) and men the remaining 49.8% ($n = 113$). Regarding marital status, the majority was married or was living in cohabitation (68.3%). In addition, most of the participants were also employed (82.8%) in the tertiary sector of activities (commerce, tourism, transportation, and financial activities) (53.7%) and, particularly, in the 7th professional group (qualified industry and construction workers and craftsmans) (38.8%). The professional activities were assigned accordingly to the Portuguese Professions Classification (National Institute of Statistics [NIS], 2010).

Participants were also inquired about clinical and health issues and 20.3% reported sleep problems, 9.7% had received treatment for these problems, and 11% of the respondents mentioned taking medication for sleep problems.

2.2 Measures

For the present study, there were considered only two measures of the four used by Alves [38], namely the Sociodemographic and Clinical Data Sheet and the Glasgow Sleep Effort Scale (GSES).

2.2.1 Sociodemographic and Clinical Data Sheet

This questionnaire constitutes a brief set of sociodemographic and clinical questions to describe the sample. Regarding sociodemographic information, participants responded questions about their age, sex, marital status (single/married or cohabitation/separated/divorced/widower), education level (corresponding to the Portuguese education system) (basic education—1st cycle/2nd cycle/3rd cycle/secondary school/higher education), actual professional situation (employed/non-employed/sick leave/retired), and profession. The clinical part of the questionnaire comprised questions about the presence of sleep problems (yes/no), treatment of sleep problems (yes/no), about taking medication or sleep supplements (yes/no), diagnosis of physical or mental disease (yes/no) and, finally, a question about taking other types of medication (yes/no) [38].

2.2.2 Glasgow Sleep Effort Scale (GSES; Broomfield & Espie, 2005; European Portuguese Adaptation by Meia-Via et al., 2012)

The GSES is a measure of one's effortful preoccupation with sleep. It is constituted by seven items that are referred to past week's sleep pattern and are assessed on a 3-point Likert scale (0 = not all, 1 = to some extent, and 2 = very much). Total scores are obtained by the sum of the score of each item, in a possible total of 14 points. Higher score denotes greater sleep effort [7]. The internal consistency value for the current sample was 0.82 (cf. Introduction).

2.3 Procedures

The procedure here described is the one adopted by Alves [38]. First, it was requested permission to the authors responsible for the GSES European Portuguese version. According to the population characterization presented in NIS [39], data were collected in several locations (e.g., public streets, workplaces, or through family relatives or acquaintances) to obtain the most representative sample of the Portuguese population possible. The instruments were applied individually or in small groups, always ensuring the adequate conditions for its application. Participants were informed on the study goals and gave their informed consent. The study was approved by the Department and University to which the last author belongs. In addition, the study was performed in line with the principles of the Declaration of Helsinki.

2.4 Data Analysis

Descriptive statistics of the sample, such as means, standard deviations, frequencies, minimum and maximum scores, as well as inference tests and operative properties,

were conducted using IBM SPSS Statistics v.27 for Windows. Considering a sample with $n > 30$, the Central Limit Theorem (CLT) and observed skewness and kurtosis values within -2 and 2 range, it may be assumed that the dataset follows approximately a normal distribution [40, 41]. However, a more robust estimation method—bootstrapping—was used to reduce bias in the analysis. There were used 1000 bootstrap samples by default, and there were determined Bias corrected accelerated (BCa) bootstrapped 95% confidence intervals (CIs) [40].

A first analysis aimed to evaluate significant differences between men and women and between groups with and without sleep problems was performed with independent samples t tests. Effect sizes were considered small, moderate, and large for Cohen's d of 0.2, 0.5, and 0.8, respectively [40]. All analyses and calculations were performed considering a level of significance at $p < 0.05$.

To conduct CFA analysis and examine underlying factorial structure of the GSES and respective goodness-of-fit indices, JASP 0.14.0.0 software was used. CFA was carried out using the Robust Diagonally Weighted Least Squares (RDWLS) estimator, which accounts for polychoric correlation matrices of Likert-type scales and does not consider the normality assumption of data [42]. To determine model adjustment, multiple fit indices such as the Chi-square test (χ^2), the Chi-Square Critical Ratio (χ^2/df), the Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), and the Root-Mean-Square Error of Approximation (RMSEA) were calculated. The model is considered adequate to the data if χ^2 is not significant, that is, $\chi^2 > 0.05$ [43] and $\chi^2/df < 5$ (ideally < 3) [44]; goodness-of-fit indices GFI and TLI with values closer to 1.0 are indicative of better fit, with minimum preferred values of 0.90 (ideal values of 0.95) [44, 45]. The $RMSEA \leq 0.08$ is considered preferable (ideally ≤ 0.06) [43]. Pertaining to local adjustment of the model, the standardized factors' loadings (λ) must be significant ($p < 0.05$) and equal or superior to 0.50 (ideally ≥ 0.7) [44, 46]. In addition, it was also computed the *Average Variance Extracted* (AVE) to examine convergent validity of the measure as well as the composite reliability (CR), which is an alternative parameter to Cronbach's α . Adequate values of CR should be ≥ 0.70 and of AVE should be ≥ 0.50 [46].

Beyond CFA, an additional analysis of the operative properties of the GSES was conducted to examine clinical accuracy examination and identification of the optimal cut-off point. Sensitivity and specificity of the scale and the Area Under the Curve (AUC) were performed using the Receiver-Operative Characteristic (ROC) curve and the Youden's index (J). The latter was computed to suggest an optimal cut-off point. For the identification of the participants correctly identified as having “insomnia” (continuous variable) and having/not having sleep problems (dichotomic variable),

the Positive Predictive Value (PPV) and the Negative Predictive Value (NPV) were also calculated.

The IRTPRO 4.2 (Item Response Theory for Patient-Reported Outcomes) student version was used for IRT analysis. The Graded Response Model (GRM) for polytomous items was considered due to its wide utilization in IRT models for ordered Likert rating scales [26]. Two parameters' models (2PL) account for the estimation of discrimination and location along the continuum of values, which is conceived as a “z score”. Difficulty values (b) normally vary from -3 (easy items) to 3 (difficult items). Moreover, discrimination parameter (a) ranges between 0 (no discrimination) and 3 (perfect discrimination). Values comprehended between 0.6 and 1.8 were considered appropriated [24] and for some authors, values above 1.7 are considered very high [47]. $S\text{-}\chi^2$ indicator was determined to evaluate model fit with significance levels at 0.01 . Statistically significant items were considered not adjusted to the model.

Network analysis was conducted in R [48]. Regularized partial correlations were computed to estimate the network. For the purpose, EBICglasso regularization was chosen—a Gaussian graphical model which uses Least Absolute Shrinkage and Selection Operator (LASSO) with Extended Bayesian Information Criterion (EBIC) model selection [33]. This estimation procedure is useful to control spurious edges due to sampling error and minimize the probability of type-I errors [49]. In the estimated network, nodes correspond to the GSES items and edges to its correlations. Since data are ordinal (a Likert-type variable), a polychoric correlation matrix among the seven items was considered as input [33, 50]. In this case, edge weights constitute partial correlation coefficients [50]. Zero-valued correlations are indicative of independent relationships between the items. Network estimation and visualization were possible using the R -packages “bootnet”, which comprises the EBICglasso estimator (Epskamp et al., 2017) and “qgraph” [51]. Centrality indices were calculated for each item in the network, namely *strength*, *closeness*, *betweenness* and *expected influence* and were computed as Z -scores. The most central nodes were topographically displayed at the centre of the network using the Frucherman–Reigold algorithm [32].

To explore network accuracy and stability for small sample sizes, an analysis of bootstrap type was conducted using the R “bootnet” package. “Bootstrapping involves repeatedly estimating a model under sampled or simulated data and estimating the statistic of interest.” [33] (p. 5). To examine the accuracy and stability of the edge weights, non-parametric bootstrapping was used, which is recommended for ordinal data analysis. The *case-dropping bootstrap* procedure was chosen to assess the centrality indices stability [33, 52]. The correlation stability (CS) coefficient for centrality indices was determined and considered adequate if above 0.25 (preferable if > 0.50) [33]. Difference tests by

bootstrapping were determined to examine significant differences between the edge weights and node centrality [52].

3 Results

3.1 Descriptive Statistics

GSES total mean score was 3.03 ($SD = 2.95$) (min = 0 ; Max = 12). Maximum coefficient values were 1.16 for skewness and 1.06 for kurtosis, considering the sex variable. Regarding the “sleep problem” variable, maximum coefficient values were 1.29 for skewness and 1.95 for kurtosis.

3.2 Differences Between Men and Women and Individuals with and Without “Sleep Problems”

A Student's t test for independent samples with bootstrapping indicated that the GSES scores were significantly higher in women ($M = 3.57$, $SD = 3.25$) relatively to men ($M = 2.49$; $SD = 2.50$). This difference achieved a large effect size: $t(211.76) = -2.82$, $p = 0.005$, Cohen's $d = 2.90$, BCa bootstrapped $95\%CI = [-1.89$ to $-0.37]$ (Field, 2017). Relatively to sleep problems, the difference between groups was significant, achieving a large effect size: $t(57.89) = 8.49$, $p = 0.001$, Cohen's $d = 2.44$, Bca bootstrapped $95\%CI = [3.16$ – $5.02]$ [40]. The “sleep problems” group ($M = 6.30$, $SD = 3.08$) presented higher GSES scores than the “without sleep problems” group ($M = 2.20$, $SD = 2.26$).

3.3 CFA

A one-factor model—sleep effort—was tested. The overall fit was not significant $\chi^2(14) = 11.85$, $p = 0.619$, suggesting a good fit between data and the proposed model. Other fit indices: $\chi^2/df = 0.85$; $CFI = 1.00$, $TLI = 1.00$ and $RMSEA = 0.00$ [$90\% IC = 0.00$ – 0.055], were considered excellent. Considering local model adjustment (cf. Fig. 1), all factor loadings were statistically significant ($p \leq 0.001$) ranging from 0.54 (Item 2) to 0.76 (Item 7), which is indicative of good local adjustment. According to the recommendations of several authors [44–45], the one-factor model for GSES in the current sample was considered well-adjusted to data.

In terms of convergent validity, it was obtained an $AVE = 0.43$ which is below the recommended value of 0.50 . *Composite reliability (CR)* was 0.84 , which is considered a good value, surpassing the recommended value of 0.70 .

3.4 GSES Operative Properties

The obtained ROC curve is displayed in Figure S1 (see Supplementary Material). The optimal cut-off point

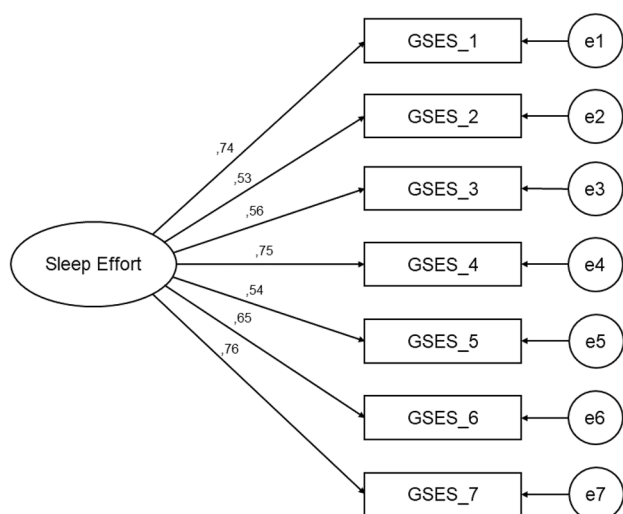


Fig. 1 Confirmatory factor analysis of the GSES. *Note.* Standardized coefficients are displayed. *Labels:* GSES_1: “I make too much of an effort to fall asleep, when it should happen naturally”; GSES_2: “I feel that I should be able to control my sleep”; GSES_3: “At night, I put off going to bed for fear of not being able to sleep”; GSES_4: “If I can’t fall asleep I get worried”; GSES_5: “I feel like I’m not very good at sleeping”; GSES_6: “Before going to bed I get anxious about my sleep”; GSES_7: “I worry about the consequences of not sleeping”

score according to the Youden’s index (*J*) in identifying an individual with “sleep problems” was 4 with 78.3% of sensitivity and 77.3% of specificity (*AUC* = 0.86, 95%CI = [0.80–0.92], *SE* = 0.03; *p* < 0.005). The proportion of individuals correctly identified by the measure as having “sleep problems” (PPV) was of 78.3%, 95%CI = [65.39–87.27] and the proportion of participants correctly identified as “not having sleep problems” and who do not have them (NPV) was of 77.4%, 95%CI = [73.39–80.87].

3.5 IRT Analysis

Examining the assumptions of IRT analysis, unidimensionality was assessed with CFA (cf. CFA section), and a single factor was identified—sleep effort—with good goodness-of-fit indices. Considering the local independence assumption, LD statistics were all values under the threshold of |10| used to indicate violation of this principle (*Min* = −0.5; *Max* = 2.8) [53], which is evidence for local independency. Monotonicity was also evidenced by the direct observation of the Test Information Curve (cf. Figure S2). There was no significant (DIF) between the groups with and without “sleep problems” (*p* > 0.01), with *p* values minimum of 0.12 and maximum of 0.75. Items measured information equally between both groups (cf. Figure S3).

All items performed adequately in the GRM analysis for polytomous items using an alpha threshold of 0.01. In this regard, χ^2 *P* values ranged between 0.23 and 0.84 (*M* = 0.58). Nevertheless, as observed by the *p* values, items performed differently. Parameters of discrimination (*a*) and location (*b*₁ and *b*₂) are presented in Table 1. All item discrimination values (*Min* = 1.36; *Max* = 3.37) ranged from 0.60 to 1.80 [24] being considered appropriate. Items 1, 3, 4, 6, and 7 were considered items with very high discrimination, with *a* > 1.7 [47]. Item 3 had the higher discrimination parameter (*a* = 3.37), followed by item 4 (*a* = 3.18) and item 6 (*a* = 3.02). Item 5 had the lowest discrimination parameter (*a* = 1.36), which is explained by the high probability of not endorsing this item (69.2% of the participants responded negatively to this item). Values of *b*₁ and *b*₂ ranged between −0.23 and 1.42 and 1.19 and 2.61, respectively, which indicates satisfactory coverage of the latent trait.

A visual representation of each item precision is presented in Fig. 2. Each individual plot shows the contribution of each item to the scale in terms of given information across a range of theta values.

As observed, item 3 can be considered the most informative item, however, in a restrict theta value range.

Table 1 GSES item description and corresponding parameter estimates and fit statistics

Item	Parameter estimates			S- χ^2
	<i>a</i>	<i>b</i> ₁	<i>b</i> ₂	χ^2
1. I make too much of an effort to fall asleep, when it should happen naturally	2.55	0.48	1.66	8.06
2. I feel that I should be able to control my sleep	1.40	0.23	2.61	11.81
3. At night, I put off going to bed for fear of not being able to sleep	3.37	1.42	2.10	11.22
4. If I can’t fall asleep I get worried	3.18	−0.11	1.64	7.65
5. I feel like I’m not very good at sleeping	1.36	0.80	2.21	20.83
6. Before going to bed I get anxious about my sleep	3.02	1.09	2.30	5.34
7. I worry about the consequences of not sleeping	2.84	−0.23	1.19	8.16

GSES Glasgow Sleep Effort Scale

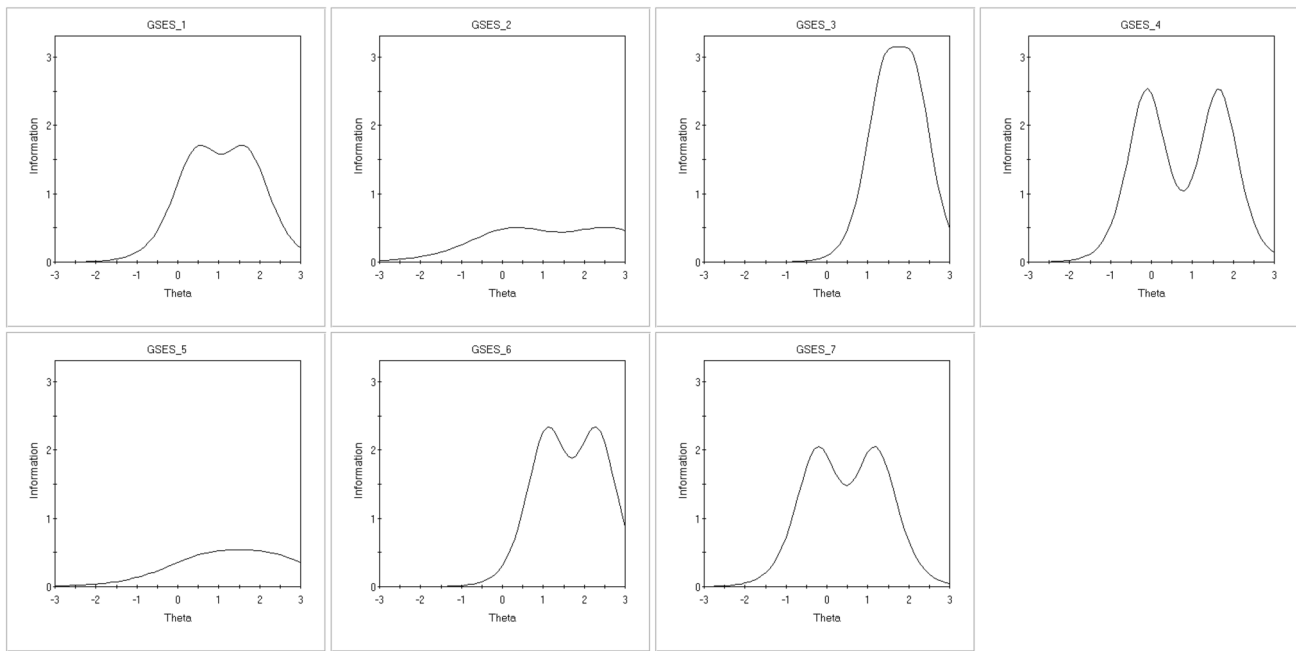


Fig. 2 IRT-based Item Information Curves for the European Portuguese version of the GSES ($N=227$). *Note:* Values of Θ are displayed in the horizontal axes, ranging from -3 to 3 , whereas item information is displayed in the y-axes, ranging from 0 to 3.5 , corresponding to the precision of each item measuring the construct. *Labels:* GSES_1: “I make too much of an effort to fall asleep, when it should

happen naturally”; GSES_2: “I feel that I should be able to control my sleep”; GSES_3: “At night, I put off going to bed for fear of not being able to sleep”; GSES_4: “If I can’t fall asleep I get worried”; GSES_5: “I feel like I’m not very good at sleeping”; GSES_6: “Before going to bed I get anxious about my sleep”; GSES_7: “I worry about the consequences of not sleeping”

Particularly, this item showed the most precise information when considering higher latent trait—sleep effort—levels. The next most informative items across the theta *continuum* were item 4, 6 and 7. Items 5 and 2 can be identified as the less informative items, despite providing some information on the evaluated construct. Items 1, 2, 4, 6, and 7 presented two theta intervals where most information on the construct is given. The majority of the items provided less information for negative values of theta. On the other hand, all items were informative for positive levels of the latent trait. Considering, for example, item 6, the most precise information was mainly present in positive theta values. The information graphic for item 4 displayed an information peak around the average theta level ($\Theta = 0$) and another in the positive range of values.

Concerning total scale information, the Total Information Curve plot with the corresponding standard errors is displayed in Figure S4. Higher information scores were concentrated between a theta interval of 1 and 2, which represents relatively high sleep effort levels. As expected, as the latent trait value increases, expected scores increase as well.

3.6 Network Analysis

Figure 3 displays the resulting graphical representation of the network structure pertaining to the GSES items. This network comprised 19 non-zero edges out of 21 possible edges (mean weight 0.13). Items 4 and 7 presented the strongest edge weight (0.48). Similarly, items 1 and 5 (0.35) and items 3 and 6 (0.34) were strongly correlated. Item 1 was displayed at the centre of the network and encompasses the largest number of connections with other items. Nonetheless, it is also worth noting that several nodes from different clusters are associated and all correlations were positive.

Centrality indices (*betweenness*, *closeness*, *strength* and *expected influence*) of each item are visually presented in Fig. 4. Item 1, directly related to sleep effort, was highlighted as the most central and important node of the network, with the highest values of *betweenness* (1.96), *closeness* (1.48), *strength* (1.34), and *expected influence* (1.34). On the other hand, item 2 appeared as the item with lower centrality indices.

Nevertheless, the observed differences in the computed parameters (e.g., edge weight and centrality indices) are justified when considered the accuracy of the network

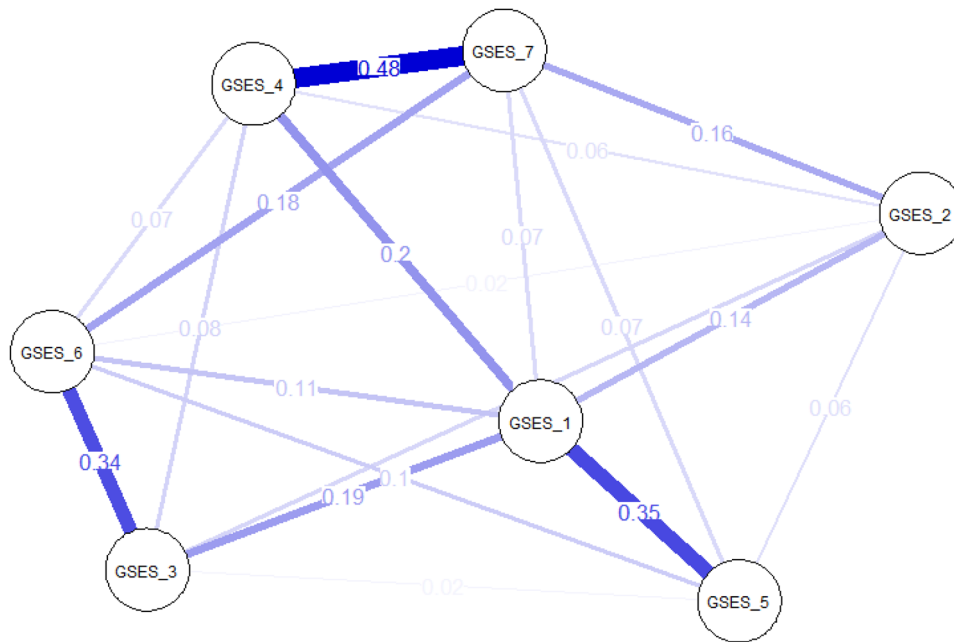


Fig. 3 Estimated Network Structure of the European Portuguese version of GSES ($N=227$). *Note:* Blue lines denote positive associations. If there were negative correlations, they would be represented in red colour. Thicker and brighter edges are indicative of stronger associations. *Labels:* GSES_1: “I make too much of an effort to fall asleep, when it should happen naturally”; GSES_2: “I feel that

I should be able to control my sleep”; GSES_3: “At night, I put off going to bed for fear of not being able to sleep”; GSES_4: “If I can’t fall asleep I get worried”; GSES_5: “I feel like I’m not very good at sleeping”; GSES_6: “Before going to bed I get anxious about my sleep”; GSES_7: “I worry about the consequences of not sleeping”

structure [edges] as well as the stability of centrality indices. The large non-parametric bootstrapped 95% confidence intervals (CIs) using 2500 bootstrap samples for the network edges were indicative of significant variability in edge weight estimation (cf. Figure S5). Thus, the interpretation of the order of the edge-weight estimates must be made with caution. Network structures that are relatively unstable do not need to be discharged. Instead, those are interpreted in the presence of some problematic issues (e.g., estimation problems due to a small N size or equally strong edges) [33, 52].

Pertaining to centrality stability (cf. Figure S6), the CS-coefficient indicated that *expected influence* ($CS(\text{cor}=0.7)=0.52$) performed better than *strength* ($CS(\text{cor}=0.7)=0.44$), *closeness* ($CS(\text{cor}=0.7)=0.28$) and *betweenness* ($CS(\text{cor}=0.7)=0.13$). All centrality indices, excepting *expected influence*, were not sufficiently stable, since did not meet the recommended cut-off value of 0.50. *Strength* and *closeness* were considered adequate ($CS(\text{cor}=0.7)>0.25$) and *betweenness* did not reach the minimum cut-off of 0.25 as suggested by Epskamp et al. [33]. Only *expected influence* was a stable index for meaningful inference.

Finally, the bootstrapped test for significant differences was conducted to compare edge weights and centrality indices (cf. Figures S7 and S8).

4 Discussion

The present study had as objective the study of psychometric properties of the GSES using three different psychometric approaches. For this purpose, we examined the factor structure of the GSES, performed IRT analysis to identify the items that provided the greatest amount of information related to the latent variable (sleep effort) and identified the item network structure, highlighting its most central nodes. This is the first published study to the best of our knowledge converging the mentioned psychometric approaches in the evaluation of a test structure in the sleep medicine field [18].

The CFA suggests a unidimensional factor structure and local independence, which are assumptions for IRT analysis. All items load significantly on a single factor—sleep effort. This is indication for a model with good local adjustment and an overall good fit between data and the proposed model with excellent fit indices. The measure was proved to be reliable, with $CR=0.84$. Item 7 (0.76) and item 4 (0.75) are related to worry with the consequences of not sleeping and worry of not being able to fall asleep, respectively, and item 1 (0.74) is related to increased effort to fall asleep, present the greatest factor loadings. In the European Portuguese version of the GSES, a cut-off point was not determined [9], so we decided to examine this topic in a more diverse sample, establishing

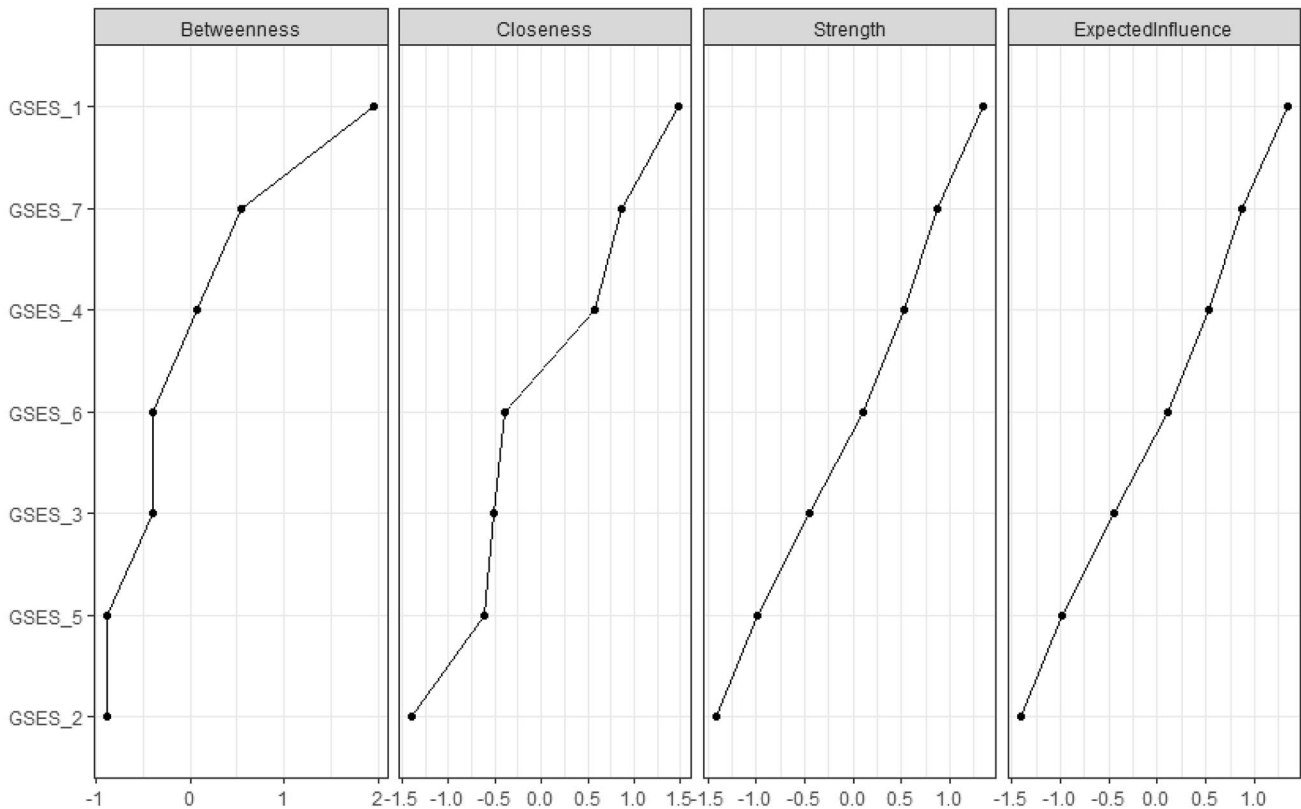


Fig. 4 Standardized scores regarding centrality indices for the GSES items. *Note:* Centrality indices are presented as Z-scores. *Labels:* GSES_1: “I make too much of an effort to fall asleep, when it should happen naturally”; GSES_2: “I feel that I should be able to control my sleep”; GSES_3: “At night, I put off going to bed for fear of

not being able to sleep”; GSES_4: “If I can’t fall asleep I get worried”; GSES_5: “I feel like I’m not very good at sleeping”; GSES_6: “Before going to bed I get anxious about my sleep”; GSES_7: “I worry about the consequences of not sleeping”

a 4-point cut-off for this scale. This is an important contribution due to the pathognomonic characteristics of this measure and its wide utilization in clinical practice.

Concerning IRT, the results indicate that the items performed differently in terms of the given information on sleep effort. All items present good discrimination parameters, whereby items 1, 3, 4, 6, and 7 are considered very discriminative ($a > 1.7$). This is a conclusion that goes in the same direction that the factor loadings previously identified for items 1, 4, and 7, with higher factor loadings for these items. Item 3 ($a = 3.18$), related to the avoiding behaviour of putting off going to bed due to fear of not being able to sleep, has the greatest item precision for higher levels of sleep effort. Item 4, related to worry of not being able to fall asleep, and item 6, related to sleep anxiety before going to bed, are the following most discriminative items, with two differential levels of theta where the higher item precision is reached. Overall based on our results, the GSES was proved to be a good measure in individuals with higher values of the latent variable—sleep effort—that is, those who endorse moderate-to-severe symptomatology.

Moreover, items 2 ($a = 1.40$), related to self-abilities to control one’s sleep, and 5 ($a = 1.36$), related to sleep self-efficacy, present poor performance, being the least discriminative items of the scale and also giving the lowest amount of information. This is a congruent conclusion with the low factor loading in CFA for mentioned items. The results demonstrate that, statistically and in the current sample, those items may be not considered as much useful as the remaining, which does not mean that the same pattern of response will be identified for other samples, namely clinical ones. Concerning groups, a DIF analysis demonstrates a nonsignificant difference between the groups with and without “sleep problems”, suggesting no influence of other dimensions beyond the main construct in the response to the items. The GSES seems to be an equivalent measure in both groups [24].

Concerning NA, the most central node of the network is item 1 ($strength = 1.34$), directly associated with effort to fall asleep. It is also one of the items with the greatest factor loading on sleep effort. Then, item 7 and item 4 appear as significant central nodes, a result that is in line with was found for the items with higher factor loading on the latent

variable. Considering all psychometric approaches, item 4, directly related to the worry of not being able to fall asleep, loads significantly on the latent variable, is very discriminative and one of the most important nodes of the estimated network. Contrary to these findings, item 2, related to self-abilities to control one's sleep presents the lowest factor loading, is the second least discriminative item and presents the lowest centrality indices. Additionally, according to the network structure, item 4 is strongly correlated to item 7, and these are related to worry of not falling sleep and the consequences of not sleeping, respectively. Likewise, items 1 and 5 are strongly correlated and associated to effort to fall asleep and abilities in falling asleep, respectively. Also, there is a strong correlation between item 3 and 6, that refer to avoidance behaviours in getting to bed for fear of not being able to sleep and the anxiety about one's sleep, respectively.

Interestingly, the evaluation of the GSES with the three psychometric approaches revealed that the items assessing worry in not being able to fall asleep (i4) and the consequences of not sleeping (i7), as well as the one related to the increased effort to fall asleep (i1), are crucial items regarding clinical practice. Patients with insomnia are actively focused on their sleep difficulties and commonly present performance anxiety about sleeping with dysfunctional thoughts. Insomniac patients often exhibit excessive and conscious effort and intention when lying in bed to fall asleep, which is associated with cognitive and physiological arousal. However, they may be able to fall asleep in contexts outside their bedrooms and without the intention to sleep [2]. Thereby, in clinical practice, it may be useful to use the GSES in the evaluation of sleep behaviours and sleep hygiene, considering cognitive-behavioural therapy for insomnia (CBT-I) [54]. Accordingly, CBT-I-related techniques (e.g., stimulus control and sleep restriction treatments) may be useful to modify the Attention–Intention–Effort pathway [5, 6], contributing to the overcoming of sleep inhibitory mechanisms and reestablishment of adaptative “setting conditions” for the reinstallation of default sleep [5].

Despite distinguish characteristics between the mentioned psychometric methods, network approaches must be considered as a complementary in the application of CTT and IRT [36]. It allows a more complete assessment of measures, and the reduction of several limitations in each of the approaches [55]. NA is developed as a promising approach in psychometrics, psychological evaluation, and more specifically in research on test development and sophistication. It becomes interesting the use of NA in the comprehension of psychological phenomena and intervention in psychopathology, considering individuals with their idiosyncrasies [34]. Previous approaches, e.g., IRT, posit that symptoms covariance is a consequence of an underlying latent variable—a common latent cause [32, 34]. However, the latter is related to a few limitations. According to NA approaches in

psychopathology, symptoms are not a reflection of an underlying mental disorder but instead elements of its constitution [32]. NA is a major contribution in the study of comorbidity with bridge symptoms analysis. Particularly, identifying the most central nodes and its interactions allows the activation/deactivation of the links in the network structure and tailor intervention in the most influential nodes [56].

This study is a contribution to the utilization and development of sophisticated self-report measures within behavioural sleep medicine and it enables the understanding of sleep effort phenomenon. Nevertheless, and despite an approximately representative sample of the Portuguese population comprising a diverse and heterogeneous sample described according to CENSOS 2011 [39], this constitutes a community sample with relatively small size. Therefore, any generalization to well-defined clinical groups or the Portuguese population must be made with caution. Despite the number of participants, our analyses are justified and robust. In CFA, in terms of absolute numbers, the acceptable minimum sample size is 50 participants for measurement models, varying up to 400 participants. In terms of proportion/ratio, there are recommended 5 participants up to 20 or more participants per item. Our sample comprises 227 participants and nearly 32 participants per item, and with the mentioned reasons, our analysis is robust [23]. Considering IRT, despite a relatively small sample size, and the requirement comprises at least 250 responses (ideally 500) [57], the GSES items were considered very discriminative. Moreover, small samples in IRT analysis are quite common in other studies [58]. In NA and psychological networks, centralities indices such as strength are parameters estimated from sample data, which are directly related to the sample size [30, 33]. Furthermore, due to the novelty of this approach, there is no sample dimension reference method in this type of analysis to obtain accurate and stable edge weights as well as centrality estimates [59].

In the sleep medicine field, non-clinical samples are often used due to the accessibility in data collection [60], and the GSES seems to be a relevant measure in primary health care. In further studies, it would be beneficial to collect data from a larger and representative sample of the Portuguese population, with an update of each of the categories, namely sex, education level, activity sector, and professional group. It would be also relevant to explore a larger number of clinical samples beyond patients with insomnia. Finally, we posit that is essential for the evaluation of measures used in our clinical and research practice (in sleep medicine and other domains) concerning a complementary integration of the psychometric approaches.

Due to the standardized utilization of sleep measures in clinical assessment and therapy as a diagnostic complementary exam, we believe that the conclusions obtained with the present study can establish a starting point in the

consideration of more advanced psychometric approaches in test sophistication. The promising study's results suggest that the GSES seems to be a good example of a sleep effort measure for the Portuguese population with higher latent trait—sleep effort—values, that is, with moderate-to-severe symptomatology. Regarding practical implications of this study, we underline the development of shorter but still reliable and valid self-report measures to bridge the concerning with the overloading of patients with extensive research protocols and clinical evaluations [18, 37]. Ultimately, we suggest that it is possible to identify potential core items of the scale in assessment and as intervention tools for insomnia, without compromising its discriminative power. To this end, we may consider, for example, item factor loadings in the sleep effort construct (for CFA), item parameter levels and item and test information curves (for IRT analysis), and node centrality (for NA) combined.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s41782-023-00229-4>.

Acknowledgements The authors would like to express their gratitude to Juliana Alves for collecting data.

Funding Open access funding provided by FCTIFCCN (b-on).

Data availability The dataset analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interest The authors declare no conflicts of interest.

Ethics Approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the University of Aveiro.

Research Involving Human Participants and/or Animals All procedures performed in this study were in accordance with the ethical standards of institutional research committee and with the 1964 Helsinki Declaration and its later amendments.

Informed Consent All participants provided signed written informed consent before participation in this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Lichstein KL, Taylor DJ, McCrae CS, Petrov ME. Insomnia: epidemiology and risk factors. In: Kryge M, Roth T, Dement WC, editors. Principles and practice of sleep medicine. Amsterdam: Elsevier; 2016.
- American Academy of Sleep Medicine (AASM). International classification of sleep disorders: diagnostic and coding manual. 3rd ed. Rochester: American Academy of Sleep Medicine; 2014.
- American Psychiatric Association (APA). DSM-5: Diagnostic and statistical manual of mental disorders. Arlington: American Psychiatric Association Publishing; 2014.
- Marques DR, Gomes AA, Clemente V, Santos JM, Castelo-Branco M. Hyperarousal and failure to inhibit wakefulness in primary insomnia: “Birds of a feather”? Sleep Biol Rhythms. 2015. <https://doi.org/10.1111/sbr.12115>.
- Espie CA. Insomnia: conceptual issues in the development, persistence, and treatment of sleep disorder in adults. Annu Rev Psychol. 2002. <https://doi.org/10.1146/annurev.psych.53.100901.135243>.
- Espie CA, Broomfield NM, MacMahon KM, Macphee LM, Taylor LM. The attention–intention–effort pathway in the development of psychophysiological insomnia: a theoretical review. Sleep Med Rev. 2006. <https://doi.org/10.1016/j.smrv.2006.03.002>.
- Broomfield NM, Espie CA. Towards a valid, reliable measure of sleep effort. J Sleep Res. 2005. <https://doi.org/10.1111/j.1365-2869.2005.00481.x>.
- Kohn L, Espie CA. Sensitivity and specificity of measures of the insomnia experience: a comparative study of psychophysiological insomnia, insomnia associated with mental disorder and good sleepers. Sleep. 2005. <https://doi.org/10.1093/sleep/28.1.104>.
- Meia-Via MS, Marques DR, Espie CA, da Silva CF, Allen Gomes A. Psychometric properties of Glasgow Sleep Effort Scale in Portuguese language. Psychol Assess. 2016. <https://doi.org/10.1037/pas0000178>.
- Crawford MR, Chirinos DA, Iurcotta T, Edinger JD, Wyatt JK, Manber R, Ong JC. Characterization of patients who present with insomnia: is there room for a symptom cluster-based approach? J Clin Sleep Med. 2017. <https://doi.org/10.5664/jcsm.6666>.
- Espie CA, Barrie LM, Forgan GS. Comparative investigation of the psychophysiological and idiopathic insomnia disorder phenotypes: psychologic characteristics, patients' perspectives, and implications for clinical management. Sleep. 2012. <https://doi.org/10.5665/sleep.1702>.
- Hertenstein E, Nissen C, Riemann D, Feige B, Baglioni C, Spiegelhalter K. The exploratory power of sleep effort, dysfunctional beliefs and arousal for insomnia severity and polysomnography-determined sleep. J Sleep Res. 2015. <https://doi.org/10.1111/jsr.12293>.
- Spiegelhalter K, Regen W, Baglioni C, Klöppel S, Abdulkadir A, Hennig J, Nissen C, Riemann D, Feige B. Insomnia does not appear to be associated with substantial structural brain changes. Sleep. 2013. <https://doi.org/10.5665/sleep.2638>.
- Malaffo M. The quarter of an hour rule: A simplified cognitive–behavioural intervention for insomnia improves sleep [Doctoral dissertation, University of Glasgow]. Enlighten: Theses; 2006. <https://theses.gla.ac.uk/id/eprint/1529>
- Kyle SD, Morgan K, Spiegelhalter K, Espie CA. No pain, no gain: An exploratory within-subjects mixed-methods evaluation of the patient experience of sleep restriction therapy (SRT) for insomnia. Sleep Med. 2011. <https://doi.org/10.1016/j.sleep.2011.03.016>.
- Fairholme CP, Manber R. Safety behaviors and sleep effort predict sleep disturbance and fatigue in an outpatient sample

- with anxiety and depressive disorders. *J Psychosom Res.* 2014. <https://doi.org/10.1016/j.jpsychores.2014.01.001>.
17. Bei B, Asarnow LD, Krystal A, Edinger JD, Buysse DJ, Manber R. Treating insomnia in depression: insomnia related factors predict long-term depression trajectories. *J Consult Clin Psychol.* 2018. <https://doi.org/10.1037/ccp0000282>.
 18. Dias SF, Marques DR. “Life” beyond classical test theory: some considerations on using complementary psychometric approaches in sleep medicine. *Sleep Med.* 2021. <https://doi.org/10.1016/j.sleep.2020.11.012>.
 19. Filho NH, Zanon C. Questões básicas sobre mensuração. In: Hutz C, Bandeira DR, Trentini CM, editors. *Psicometria*. Porto Alegre: Artmed; 2015. p. 23–39.
 20. Timmerman ME, Lorenzo-Seva U, Ceulemans E. The number of factors problem. In: Irwing P, Booth T, Hughes DJ, editors. *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*. Chichester: Wiley; 2018. p. 305–24.
 21. Dancey CP, Reidy J. Introduction to factor analysis. In: Dancey CP, Reidy J, editors. *Statistics without maths for psychology*. Harlow: Pearson; 2017. p. 446–80.
 22. Pacico JC, Hutz CS. Validade. In: Hutz C, Bandeira DR, Trentini CM, editors. *Psicometria*. Porto Alegre: Artmed; 2015. p. 71–83.
 23. Furr RM, Bacharach VR. *Psychometrics: An introduction*. 2nd ed. Thousand Oaks: Sage; 2013.
 24. Nakano TC, Primi R, Nunes CHSS. Análise de itens e teoria de resposta ao item (IRT). In: Hutz C, Bandeira DR, Trentini CM, editors. *Psicometria*. Porto Alegre: Artmed; 2015. p. 97–121.
 25. Reise SP, Ainsworth AT, Haviland MG. Item response theory: fundamentals, applications, and promise in psychological research. *Curr Dir Psychol Sci.* 2005. <https://doi.org/10.1111/j.0963-7214.2005.00342.x>.
 26. Yang F, Kao ST. Item response theory for measurement validity. *Shanghai Arch Psychiatry.* 2014. <https://doi.org/10.3969/j.issn.1002-0829.2014.03.010>.
 27. Pasquali L, Primi R. Fundamentos da teoria da resposta ao item: TRI. *Avaliacao Psicol.* 2003. <https://dialnet.unirioja.es/servlet/articulo?codigo=5115864>
 28. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. *Patient.* 2014. <https://doi.org/10.1007/s40271-013-0041-0>.
 29. Fonseca-Pedrero E. Network analysis: a new way of understanding psychopathology? *Rev Psiquiatr Salud Ment.* 2017. <https://doi.org/10.1016/j.rpsmen.2017.10.005>.
 30. Hevey D. Network analysis: a brief overview and tutorial. *Health Psychol Behav Med.* 2018. <https://doi.org/10.1080/21642850.2018.1521283>.
 31. Borsboom D, Cramer AO. Network analysis: an integrative approach to the structure of psychopathology. *Annu Rev Clin Psychol.* 2013. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>.
 32. McNally RJ. Can network analysis transform psychopathology? *Behav Res Ther.* 2016. <https://doi.org/10.1016/j.brat.2016.06.006>.
 33. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: a tutorial paper. *Behav Res Methods.* 2017. <https://doi.org/10.3758/s13428-017-0862-1>.
 34. Fonseca-Pedrero E. Network analysis in psychology. *Papeles del Psicol.* 2018. <https://doi.org/10.23923/pap.psicol2018.285>.
 35. Marques DR, de Azevedo MHP. Potentialities of network analysis for sleep medicine. *J Psychosom Res.* 2018. <https://doi.org/10.1016/j.jpsychores.2018.05.019>.
 36. Schlechter P, Wilkinson PO, Knausenberger J, Wanninger K, Kamp S, Morina N, Hellmann JH. Depressive and anxiety symptoms in refugees: Insights from classical test theory, item response theory and network analysis. *Clin Psychol Psychother.* 2021. <https://doi.org/10.1002/cpp.2499>.
 37. Thakral M, Von Korff M, McCurry SM, Morin CM, Vitiello MV. ISI-3: evaluation of a brief screening tool for insomnia. *Sleep Med.* 2021. <https://doi.org/10.1016/j.sleep.2020.08.027>.
 38. Alves JS. Qualidade de sono (BaSIQS), esforço para dormir (GSES) e pensamentos antes de adormecer (GCTI) numa amostra não clínica de adultos [Sleep quality (BaSIQS), sleep effort (GSES) and thoughts before falling asleep (GCTI) in a non-clinical adult sample]. Master’s thesis, Universidade de Aveiro, Repositório Institucional da Universidade de Aveiro; 2015. <https://ria.ua.pt/handle/10773/15577>
 39. National Institute of Statistics [NIS]. *Classificação Portuguesa das Profissões 2010*. 2011. https://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=107962055&att_display=n&att_download=y. Accessed Nov 2020
 40. Field A. *Discovering statistics using IBM SPSS Statistics*. 5th ed. Los Angeles: Sage; 2005.
 41. Lomax R, Hahs-Vaughn D. *An introduction to statistical concepts*. 3rd ed. New York: Routledge; 2012.
 42. Li C. The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychol Methods.* 2016. <https://doi.org/10.1037/met000093>.
 43. Weston R, Gore PA Jr. A brief guide to structural equation modelling. *Couns Psychol.* 2006. <https://doi.org/10.1177/0011000006286345>.
 44. Brown TA. *Confirmatory factor analysis for applied research*. 2nd ed. New York: The Guilford Press; 2005.
 45. Kline RB. *Principles and practice of structural equation modeling*. 3rd ed. New York: The Guilford Press; 2005.
 46. Hair JF, Black WC, Babin BJ, Anderson RE. *Multivariate data analysis*. 8th ed. Andover: Cengage; 2019.
 47. Baker FB, Kim SH. *The basics of item response theory using R*. New York: Springer; 2017.
 48. R Core Team. *R: A language and environment for statistical computing*. Vienna: R Core Team; 2020.
 49. Hirota T, McElroy E, So R. Network analysis of internet addiction symptoms among a clinical sample of Japanese adolescents with autism spectrum disorder. *J Autism Dev Disord.* 2020. <https://doi.org/10.1007/s10803-020-04714-x>.
 50. van Zyl CJJ. A network analysis of the General Health Questionnaire. *J Health Psychol.* 2018. <https://doi.org/10.1177/1359105318810113>.
 51. Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. Qgraph: Network visualizations of relationships in psychometric data. *J Stat Softw.* 2012. <https://doi.org/10.18637/jss.v048.i04>.
 52. Burger J, Isvoranu A, Lunansky G, Haslbeck JMB, Epskamp S, Hoekstra RHA, Fried EI, Borsboom D, Blanken T. Reporting standards for psychological network analyses in cross-sectional data. *Psychol Methods.* 2022. <https://doi.org/10.1037/met0000471>.
 53. Toland MD. Practical guide to conducting an item response theory analysis. *J Early Adolesc.* 2014. <https://doi.org/10.1177/0272431613511332>.
 54. Riemann D, Baglioni C, Bassetti C, Bjorvatn BB, Groselj LD, Ellis JG, Espie CA, Garcia-Borreguero D, Gjerstad M, Gonçalves M, Hertenstein E, Jansson-Fröjmark M, Jennum PJ, Leger D, Nissen C, Parrino L, Paunio T, Pevernagie D, Verbraecken J, Weeß HG, Wichniak A, Zavalko I, Arnardottir ES, Deleau OC, Strazisar B, Zoetmulder M, Spiegelhalter K. European guideline for the diagnosis and treatment of insomnia. *J Sleep Res.* 2017. <https://doi.org/10.1111/jsr.12594>.
 55. Sartes LMA, Souza-Formigoni MLOD. Avanços na psicometria: da teoria clássica dos testes à teoria de resposta ao item. *Psicol Reflex Crit.* 2013. <https://doi.org/10.1590/S0102-7972013000200004>.

56. Machado WL, Vissoci J, Epskamp S. Análise de rede aplicada à psicometria e à avaliação psicológica. In: Hutz C, Bandeira DR, Trentinim CM, editors. *psicometria*. Porto Alegre: Artmed; 2015. p. 125–44.
57. Schultz KS, Whitney DJ, Zickar MJ. *Measurement theory in action: case studies and exercises*. 2nd ed. New York: Routledge; 2014.
58. Ventuneac A, Rendina HJ, Grov C, Mustanski B, Parsons JT. An item response theory analysis of the Sexual Compulsivity Scale and its correspondence with the Hypersexual Disorder Screening Inventory among a sample of highly sexually active gay and bisexual men. *J Sex Med*. 2015. <https://doi.org/10.1111/jsm.12783>.
59. McWilliams LA, Sarty G, Kowal J, Wilson KG. A network analysis of depressive symptoms in individuals seeking treatment for chronic pain. *Clin J Pain*. 2017. <https://doi.org/10.1097/AJP.0000000000000477>.
60. Marques DR, Gomes AA, de Azevedo MHP. Utility of studies in community-based populations. *Sleep Vigil*. 2021. <https://doi.org/10.1007/s41782-021-00135-7>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.