AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

ISSN: 1040-3590

2021, Vol. 33, No. 2, 111–121
http://dx.doi.org/10.1037/pas0000959

Brief Version of the Pittsburgh Sleep Quality Index (B-PSQI) and Measurement Invariance Across Gender and Age in a Population-Based Sample

Clara Sancho-Domingo, José Luis Carballo, and Ainhoa Coloma-Carmona
Miguel Hernández University of Elche

Daniel J. Buysse
University of Pittsburgh

The Pittsburgh Sleep Quality Index (PSQI) is the most widely used questionnaire in research and clinical practice to assess sleep quality. However, a brief version of this measure would improve its efficiency and applicability. This study aimed to develop a brief form of the PSQI and to study measurement invariance across gender and age in a nonclinical population. In total, 609 participants with a mean age of 37.3 years (standard deviation [*SD*] = 11.9) were recruited, of whom 71.8% (n = 437) were women. Participants completed online versions of the PSQI and the Insomnia Severity Index (ISI). Reliability analyses were performed to reduce the number of items, followed by validity and measurement invariance analyses for the new Brief Version of the PSQI (B-PSQI). Six questions were included in the B-PSQI out of the initial 18; the brief form had adequate internal consistency (α = .79 and ω = 0.91). Confirmatory factor analysis showed optimal fit of the B-PSQI ($\chi^2(4)$ = 22.428; $p < .01$; comparative fit index (CFI) = 0.99; normed fit index (NFI) = 0.99; Tucker-Lewis index (TLI) = 0.98; root mean squared error of approximation (RMSEA) = 0.06; standardized root mean square residual (SRMR) = 0.04), achieving partial scalar invariance across gender-same factorial structure, loadings, and thresholds in the majority of the items. Invariance across age was only achieved for model structure. Additionally, the B-PSQI yielded favorable sensitivity (75.82%) and specificity (76.99%) for classifying poor sleepers, similar to values for the full PSQI. In conclusion, the B-PSQI is a brief, reliable, and valid measure that can be used as a screening tool, allowing valid score comparisons between men and women of similar age.

Public Significance Statement
A Brief Version of the Pittsburgh Sleep Quality Index (B-PSQI) was developed to improve its efficiency and applicability. The 6-item B-PSQI is a reliable and valid tool to assess sleep quality and identify poor sleepers. The B-PSQI achieved invariance across gender, allowing valid comparisons of sleep quality between men and women of similar age. The findings highlight the efficiency of the B-PSQI and its wide potential use in assessing sleep quality.

Keywords: sleep quality, Pittsburgh Sleep Quality Index, brief form, measurement invariance

Sleep problems have emerged as a public health concern because of their negative impact on physical and mental health (Ford, Cunningham, Giles, & Croft, 2015; Jike, Itani, Watanabe, Buysse, & Kaneita, 2018; Stranges, Tigbe, Gómez-Olivé, Thorogood, & Kandala, 2012). The importance of sleep quality as a correlate of physical, mental, and cognitive health is heightened by evidence from some studies that show an increasing prevalence of sleep problems worldwide (Adams et al., 2017; Buysse, 2014; Ford et al., 2015; Zomers et al., 2017).

This article was published Online First October 29, 2020.
Clara Sancho-Domingo, José Luis Carballo, and Ainhoa Coloma-Carmona, Department of Health Psychology, Miguel Hernández University of Elche; Daniel J. Buysse, Department of Psychiatry, University of Pittsburgh.
This research received no specific grant from any funding agency, commercial entity, or not-for-profit entity. The PSQI and B-PSQI are copyrighted by the University of Pittsburgh. Their use is offered free of charge to academic researchers and clinicians. The University of Pittsburgh licenses the PSQI and B-PSQI to commercial entities for a fee. Daniel J. Buysse receives a portion of these licensing fees. The tests were reprinted or adapted with permission. Developed by Daniel J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, and D. J. Kupfer of the University of Pittsburgh using National Institute of Mental Health Funding.
Correspondence concerning this article should be addressed to José Luis Carballo, Department of Health Psychology, Miguel Hernández University of Elche, Avinguda de la Universitat d'Elx, s/n, 03202 Elche, Alicante, Spain. E-mail: jcarballo@umh.es

Although there is no consensus regarding the definition of sleep quality, it is often inferred from a combination of qualitative and quantitative sleep parameters and their impact on the waking state. Sleep parameters include subjective reports of satisfaction and quantitative estimates of sleep latency, number of awakenings, sleep duration, and sleep efficiency (percentage of time asleep while in bed). Symptoms during wakefulness generally relate to the subjective perception of restful sleep, daytime fatigue, and sleepiness (Buysse, 2014; Goelema et al., 2018; Ramlee, Sanborn, & Tang, 2017; Svetnik et al., 2020). Among the variety of methods intended to assess sleep, polysomnography is often considered the "gold standard" objective measure. However, this measure has a relatively high cost and is unfeasible when used in large samples. Moreover, polysomnography is typically conducted for one or a few nights, which may not be typical of the individual's usual pattern, and it does not capture the subjective experience of sleep. Consequently, more cost-effective methods for assessing habitual sleep are often used, such as actigraphy and self-reported measures (Corlateanu, Covantev, Botnaru, Sircu, & Nenna, 2017; Landry, Best, & Liu-Ambrose, 2015).

The Pittsburgh Sleep Quality Index (PSQI; Buysse, Reynolds, Monk, Berman, & Kupfer, 1989) is the most frequently used sleep-quality questionnaire in research and clinical practice (Mollayeva et al., 2016). This self-reported 18-item measure assesses sleep quality within the past month and includes a global score comprising seven sleep components. Because the PSQI also provides adequate sensitivity and specificity for classifying good and bad sleepers (Buysse et al., 1989), it has been often used as a screening tool (Mollayeva et al., 2016). Psychometric studies have demonstrated that the PSQI is a reliable and valid standardized measure of sleep quality in clinical and nonclinical populations (Manzar et al., 2018; Mollayeva et al., 2016). Several discrepancies have also been reported in the literature regarding the optimal PSQI factor structure or dimensionality, where one-, two-, and three-factor models have each shown adequate fit (Mollayeva et al., 2016). These types of psychometric differences could be influenced by a disparate treatment of PSQI data, such as considering data as continuous rather than ordinal, or performing analyses using the PSQI items individually versus using its seven sleep components defined in the original study (Babson, Blonigen, Boden, Drescher, & Bonn-Miller, 2012; Otte, Rand, Carpenter, Russell, & Champion, 2013).

Nevertheless, the PSQI has been used in at least 1,500 scientific studies and tested in more than 35 validation studies for diverse populations (Mollayeva et al., 2016), achieving invariance across different languages and ethnicities (Otte et al., 2013; Tomfohr, Schweizer, Dimsdale, & Loredo, 2013). PSQI invariance should also be tested across groups with expected differences in sleep quality, such as men and women or different age groups (Gadie, Shafto, Leng, & Kievit, 2017; Madrid-Valero, Martínez-Selva, do Couto, Sánchez-Romera, & Ordoñana, 2017; Mallampalli & Carter, 2014). With regard to gender, women generally report worse sleep quality than men, with a greater number of sleep disturbances, wakefulness during sleep, and longer sleep latency. These subjective reports are often inconsistent with objective sleep measures, which supports the need for testing invariance in self-reported sleep questionnaires (Mallampalli & Carter, 2014). Gender differences have also been found with polysomnography, mainly concerning the distribution of sleep stages, which can also be influenced by age (Gadie et al., 2017; Mallampalli & Carter, 2014). Across the life span, sleep duration gradually shortens and sleep efficiency decreases, particularly after age 50 (Gadie et al., 2017; Hinz et al., 2017). Older adults report greater difficulty falling asleep, more sleep disturbances, and reduced sleep efficiency compared with young and middle-aged adults (Gadie et al., 2017). Again, because of the sleep differences found between age groups, testing multiple-group invariance is important; an invariant sleep measure can ensure that the group differences are due to the sleep-quality construct rather than possible measurement errors (Byrne, 2008).

The length of the PSQI and the complexity of its scoring algorithm may limit its utility and application for some types of studies. Given the widespread use of the PSQI, a short version has been previously developed, with the purpose of reducing the burden of extensive surveys and batteries (Famodu et al., 2018). This short PSQI version was performed by testing a six-factor structure model, yielding a 13-item form. However, the reduction in the number of items from 18 to 13 is rather limited. In addition, reliability analyses were not conducted in this study but are needed to achieve an optimal reduction of items, as recommended in the literature (Widaman, Little, Preacher, & Sawalani, 2011). A substantially shorter version of the PSQI that is easier to score could provide multiple advantages, including reduced completion time, improved efficiency of data collection, and improved accuracy of responses (Galesic & Bosnjak, 2009; Rolstad, Adler, & Rydén, 2011). Other short sleep-quality measures, such as the eight-item PROMIS Sleep Disturbance and Sleep-Related Impairment (Yu et al., 2012) and the single-item Sleep Quality Scale (SQS; Snyder, Cai, DeMuro, Morrison, & Ball, 2018), provide a fast and reliable method to assess sleep quality. Unlike the PSQI, however, these sleep measures only include items with graded-response categories (e.g., never, rarely, sometimes, often, always) and do not include quantitative data (e.g., number of hours of sleep). Although such quantitative data can limit operational metrics (Yu et al., 2012), they are often useful in sleep research.

Therefore, the aim of the present study was to develop a brief form of the PSQI that includes the minimum number of items while maintaining satisfactory psychometric properties in a nonclinical population. In order to address the sleep-quality differences often found between men and women and between age groups, a second aim of this study was to analyze the measurement invariance of the brief form of the PSQI across gender and age.

Method

Participants

The estimated minimum sample size required to conduct this study was calculated according to a 10:1 ratio recommendation to perform confirmatory factor analysis (CFA; Kline, 1998). Considering that 53 parameters of variance, covariance, and regression coefficients are obtained from the 18 items of the PSQI, a ratio of 10 participants to parameters suggested a minimum of 530 individuals. We recruited a total of 665 adults of Spanish nationality, of whom 609 were used for data analyses, ensuring adequate statistical power.

Figure 1 shows a flowchart of the sample selection. From the initial sample, we first excluded individuals under age 18 ($n = 8$)

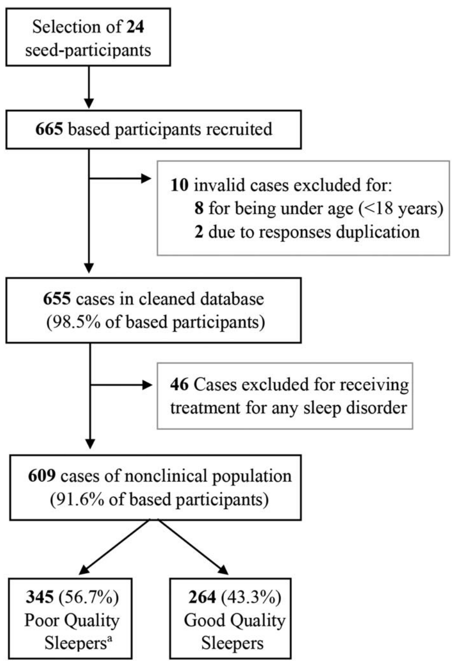

Figure 1. Flowchart of sample selection. [a] Score over 5 in the *Pittsburgh Sleep Quality Index*.

and those with duplicate responses identified through data ID code (n = 2). Additionally, 7% (n = 46) were excluded because they had been referred for medical or psychological treatment for a sleep disorder and therefore were considered to represent a clinical sample.

Participants had a mean age of 37.3 years (standard deviation [SD] = 11.9) and age range from 18 to 75. The majority were women 71.8% (n = 437), were part-time or full-time employees 70.9% (n = 432), and had university studies or higher 65.7% (n = 400). Likewise, approximately half of the sample, 49.8% (n = 303), was married or in a stable relationship. No significant differences were found between good and poor sleepers (PSQI >5; Buysse et al., 1989) for any sociodemographic variables.

Measures

Sociodemographic variables. Gender, age, level of education, employment status, and marital status were assessed by self-report. Likewise, an item was included to identify participants who were receiving treatment for any sleep disorder.

Sleep quality. We used the PSQI (Buysse et al., 1989), which assesses the quality of sleep in the past month with 18 items, with responses including self-reported times and durations or a 4-point Likert scale for frequency. The items are grouped and scored in seven sleep components: subjective sleep quality, sleep latency, sleep duration, sleep efficiency, night disturbances, use of sleeping medication, and daytime dysfunction. The sum of components scores (also coded on a Likert scale ranging from 0 to 3) results in a total global score from 0 to 21, where scores above 5 indicate poor sleep quality. The PSQI has demonstrated adequate internal consistency (Cronbach's alpha ≥ .70) and validity in diverse types of populations (Mollayeva et al., 2016).

We used the PSQI Spanish version, which has demonstrated good reliability (Cronbach's alpha = .81) and validity among the Spanish population, with an 88.6% sensitivity and a 74.2% specificity (Royuela & Macías Fernández, 1997).

Participants also completed the Insomnia Severity Index (ISI; Bastien, Vallières, & Morin, 2001), one of the most widely used measure to assess insomnia (Ibáñez, Silva, & Cauli, 2018). The ISI was included in this study in order to test the convergent validity of the brief form of the PSQI because a strong relationship between these sleep measures has been found in previous studies (Chiu, Chang, Hsieh, & Tsai, 2016; Morin, Belleville, Bélanger, & Ivers, 2011). The ISI is composed of seven items rated on a 5-point Likert scale. The questionnaire provides information about insomnia symptoms, level of satisfaction with sleep, and impact on quality of life. Global scores range from 0 to 28, where values between 0 and 7 indicate an absence of sleep problems, scores between 8 and 14 indicate subthreshold insomnia, scores between 15 and 21 indicate moderate insomnia, and scores above 22 indicate severe insomnia. This measure has demonstrated excellent validity and reliability for a community sample, with Cronbach's alpha = .9 (Morin et al., 2011). For this study, we used the ISI Spanish version (Fernandez-Mendoza et al., 2012), which has been shown to be reliable (Cronbach's alpha = .82) and valid for measuring insomnia.

Procedure

We used a web-based participant-recruitment method, which reaches a large number of people from a wide range of sociodemographic backgrounds in a cost-effective manner (Christensen et al., 2017; Kesse-Guyot et al., 2013). The Survio website platform was used to create an online survey. This survey included an initial message in which we informed participants about voluntary participation, confidentiality, and anonymity of data and asked for their informed consent. Likewise, we provided information about how to complete the survey and the estimated completion time (approximately 10 min).

We linked the completed survey to a free-access URL link in order to disseminate it online through a snowball-sampling technique. We selected a first wave of participants as chain-recruitment precursors (seed participants) from the authors' community (Wejnert & Heckathorn, 2008). Because this convenience sampling could generate a sample with similar characteristics, seed participants were selected based on differences in gender, age, level of education, and region of Spain to which they belonged, in order to ensure a varied and diverse sample. In total, 24 seed participants were recruited with diverse sociodemographic characteristics: 4 with primary education (below university level), 4 with higher education (university level or higher), 3 between 18 and 30 years old, 3 above 50 years old, 3 from the north of Spain, 4 from the south, and 3 from the center.

Once the seed participants were selected, we contacted them to inform them about the aims of this study and to ask them for their collaboration as disseminators of the online questionnaire. They were asked to share the URL link with their acquaintances and friends, for example, through instant messaging and posts in social media. To perform this task, we established a standard message to be used for all seed participants in which future participants were invited to complete the questionnaire:

> Researchers from Miguel Hernández University are conducting a sleep quality study, and they would like to count on your participation. Please complete the following questionnaire that will take you about 10 minutes to answer and share it with your contacts! Please note that your answers will be confidential and anonymous; therefore, we ask you to answer honestly to all questions. Data collected will only be used for research purposes and won't be shared with third parties. Click here for more information and to access the questionnaire.

After their completion of the questionnaire, we asked participants to share the URL link. As compensation for disseminating the questionnaire, we offered them the option of receiving their sleep-quality results via e-mail. Because compensation was not offered in the initial message, bias associated with appealing only to people interested in their sleep was reduced.

This study was reviewed and approved by the Department of Health Psychology of Miguel Hernández University of Elche. Ethical approval was not required because this was a descriptive study that did not collect personal data, and all participants gave their informed consent.

Data Analysis

To evaluate the sociodemographic characteristics, descriptive (frequencies, mean, and standard deviation) and bivariate analyses were performed using Statistical Package for the Social Sciences (SPSS) Version 24.0. Chi-square (χ^2) tests were used to analyze noncontinuous variables, and Mann–Whitney U nonparametric (Z) tests were used for continuous variables because of normality violation. The results of data analyses were interpreted while working with a 95% confidence level.

Item reduction and internal consistency. Given that PSQI data are ordinal rather than continuous, internal consistency reliability was assessed using ordinal alpha estimated with polychoric correlations. Ordinal omega was also calculated, which is an appropriate estimator when the parameters of the measure are not essentially tau equivalent, but congeneric, and when items are skewed (Gadermann, Guhn, & Zumbo, 2012; Trizano-Hermosilla & Alvarado, 2016). As a reliability criterion, a value of alpha and omega greater than 0.7 was considered acceptable (Cortina, 1993).

Reduction of the 18 PSQI items was performed sequentially, selecting items according to the following exclusion criteria: (a) presence of collinearity between items, which was examined with a polychoric correlation matrix; (b) values with corrected item-total correlation below 0.4 because values greater than 0.4 indicate better item quality (Ebel & Frisbie, 1991); and (c) maintenance or improvement of ordinal alpha when the item was excluded.

The analyses performed for item reduction and internal reliability were estimated using the *psych* package for R statistical software (Revelle, 2018).

Confirmatory factor analysis. Because the factor structure of the PSQI has shown good fit for a one-factor model (de la Vega et al., 2015; Manzar et al., 2018; Rener-Sitar, John, Bandyopadhyay, Howell, & Schiffman, 2014; Zhu, Xie, Park, & Kapella, 2018), we examined unidimensionality with CFA using the *lavaan* package for R statistical software (Rosseel, 2012).

The CFA was performed for the Brief Version of the PSQI (B-PSQI) and for the original PSQI in order to compare both versions. Because the literature indicates that PSQI psychometric properties can be analyzed using the individual items (PSQI-ITEM) and using the sleep components (PSQI-COMP), the CFA was performed for both options (Mollayeva et al., 2016).

The three examined models (B-PSQI, PSQI-ITEM, and PSQI-COMP) were estimated by robust diagonally weighted least squares (RDWLS), which provides more accurate parameter estimates and precise standard errors in skewed data compared with standard maximum-likelihood approaches (Yang-Wallentin, Jöreskog, & Luo, 2010). Furthermore, because the PSQI item sleep efficiency is derived from the item hours of sleep, we also compared the three models (B-PSQI, PSQI-COMP, and PSQI-ITEM) considering error covariance between those items (efficiency and hours of sleep), as recommended in previous studies (Ho & Fong, 2014; Raniti, Waloszek, Schwartz, Allen, & Trinder, 2018).

The models' goodness of fit was evaluated with the Satorra–Bentler scaled χ^2 statistic indicated for nonnormal data (Satorra & Bentler, 2001; Yang-Wallentin et al., 2010). Because the χ^2 statistic is sensitive to sample size and often rejects well-adjusted models (Ainur, Sayang, Jannoo, & Yap, 2017; Bentler & Bonett, 1980), we also relied on the following relative fit indexes: comparative fit index (CFI; Bentler, 1990), normed fit index (NFI; Bentler & Bonett, 1980), Tucker–Lewis index (TLI; Tucker & Lewis, 1973), root mean squared error of approximation (RMSEA; Browne & Cudeck, 1992), and standardized root mean square residual (SRMR; Chen, 2007). CFI, NFI, and TLI values ≥ 0.95 were considered as optimal fit (Hu & Bentler, 1999). RMSEA values ≤ 0.08 and SRMR values <0.05 were considered to indicate satisfactory fit (Hu & Bentler, 1999).

Measurement invariance across gender and age. Measurement invariance was tested considering the recommendations of Bowen and Masa (2015) for ordinal data. We first tested the best-fit model from CFA in men and women separately and also in three different age groups: young adults (18–34 years; $n = 292$), middle adults (35–49 years; $n = 202$), and older adults (50–75 years; $n = 115$). Next, we studied the equivalence of the B-PSQI between groups by testing three models that increase invariance stringency: (a) configural invariance model or baseline model, which implies equivalence of model form; (b) metric or weak factorial model, which refers to equivalence of loadings (λs) across groups; and (c) scalar or strong factorial model, which denotes equivalence of loadings and items thresholds (τs).

These models were estimated using RDWLS based on the Satorra–Bentler scaled χ^2 statistic. Each model was compared with its preceding model (i.e., metric model compared with configural model, scalar model compared with metric model) to study whether model fit deteriorated significantly, based on the χ^2 difference test ($\Delta\chi^2$). Again, because of χ^2 sensitivity to sample size, we also relied on models' differences in CFI (ΔCFI), RMSEA (ΔRMSEA), and SRMR (ΔSRMR) indexes where values ≤ 0.01 of ΔCFI, values ≤ 0.015 of ΔRMSEA, and values ≤ 0.03 of ΔSRMR were used to identify the most stringent model (Chen, 2007). Because items are ordered categorically, we considered acceptable ΔRMSEA values of 0.05 for metric invariance and of 0.01 for scalar invariance, as indicated by Rutkowski and Svetina (2017).

In addition, partial invariance was tested when disparity between the models was found. The forward method was used, where parameters of the noninvariant model are sequentially added or constrained to the preceding model (Jung & Yoon, 2016). We sequentially fixed nonsignificant parameters because adding this constraint would not change the fit of the model, and we retested the model until partial invariance was achieved.

To perform these analyses, we utilized the *lavaan* and *semTools* packages (Jorgensen et al., 2018; Rosseel, 2012) for R statistical software.

Convergent and concurrent validity. Convergent and concurrent validity were examined using the Spearman correlation coefficient (Fieller, Hartley, & Pearson, 1957) between B-PSQI, PSQI, and ISI. In addition, we performed receiver operating characteristic (ROC) analysis for the B-PSQI and the original PSQI to compare both measures' ability to discriminate between good and poor sleepers according to the ISI criterion cutoff of ≥ 8 (Morin et al., 2011). The area under the curve (AUC) was calculated and interpreted based on the following thresholds of discrimination power: 0.5–0.7 indicates low discrimination, 0.7–0.9 indicates moderate discrimination, and > 0.9 indicates high discrimination (Swets, 1988).

The optimal B-PSQI cutoff for screening purposes was selected by considering the global score that maximizes both sensitivity and specificity (Pintea & Moldovan, 2009). Additionally, we calculated positive predictive value (PPV), negative predictive value (NPV), and the Youden index, where higher values represent better accuracy of classification (Youden, 1950).

Results

Item Reduction and Internal Consistency

The sleep-efficiency component was treated as a single item because it is calculated with rise-time and bedtime questions. Therefore, item reduction of the PSQI was performed with the remaining 17 PSQI questions.

The polychoric correlation matrix showed that the items "How long does it usually take you to fall asleep each night?" and "Cannot get to sleep within 30 minutes" were highly correlated ($r = .71$). Therefore, because both items seemed to measure the same sleep-quality facet (sleep latency), we excluded the item "Cannot get to sleep within 30 minutes" because the quantitative response to the item "How long does it usually take you to fall asleep each night?" has clinical relevance.

Reliability analysis with the remaining 16 items yielded an initial ordinal alpha of $\alpha = .81$ and omega of $\omega = 0.85$. In total, 11 items were identified for removal: 9 because their corrected item-total correlation was below 0.4 and 2 because alpha was maintained or improved when the item was dropped. Table 1 displays the process of item reduction and the corresponding ordinal alpha and omega at each stage.

The B-PSQI included six questions (see Table 2). Because bedtime and rise time are used to calculate sleep efficiency, the six questions of the B-PSQI yield five scored items. These five items provide a global score ranging from 0 to 15, where higher scores indicate worse sleep quality. The B-PSQI had good reliability, with a polychoric ordinal alpha of $\alpha = .79$ and ordinal omega of $\omega = 0.91$. Corrected item-total scale correlations ranged from 0.51 to 0.78, demonstrating excellent discrimination index values (see Table 2).

Confirmatory Factor Analysis

Table 3 presents the results of standard and error covariance CFA for the three PSQI versions: the B-PSQI, the model calculated with the seven original components (PSQI-COMP), and the model calculated with the original items (PSQI-ITEM).

Using standard CFA, all three models were rejected statistically based on the Satorra–Bentler scaled χ^2 statistic ($p < .01$). According to the CFI, NFI, and TLI, all three models showed an adequate fit, with the B-PSQI and PSQI-COMP models exhibiting optimal goodness of fit. Only the B-PSQI model reached acceptable values for RMSEA and SRMR. Therefore, the B-PSQI model provided a satisfactory data fit for unifactorial structure, $\chi^2(5) = 39.865$, $p < .05$, CFI = 0.97, NFI = 0.98, TLI = 0.94, RMSEA = 0.08, SRMR = 0.05.

All B-PSQI standardized factor loadings were satisfactory (range $\lambda = 0.505$–0.806). By contrast, the PSQI-ITEM model included 10 variables with loadings below 0.5 (range $\lambda = 0.247$–0.770), and the PSQI-COMP model included 2 variables with loadings below 0.5 (range $\lambda = 0.232$–0.833). Based on these findings, the B-PSQI model displayed better unifactorial representation of the data than PSQI-ITEM and PSQI-COMP, with superior parsimony given the reduced number of items.

Table 1
PSQI Item Reduction and Polychoric Ordinal Alpha and Omega of Each Stage

Item exclusion criteria	Number of items excluded	Brief description of excluded items	α	ω
None	—	All items included.	.830	.862
Collinearity	1	Cannot get to sleep within 30 minutes	.814	.849
Item total correlation < .4[a]	9	Have to get up to use the bathroom	.805	.860
		Cannot breathe comfortably		
		Cough or snore loudly		
		Feel too cold		
		Feel too hot		
		Take medication to help sleep		
		Have bad dreams		
		Have pain while sleeping		
		Trouble staying awake during daytime activities		
Alpha maintained or increased when item deleted	2	Other reasons for trouble sleeping	.793	.910
		Problem in keeping up enthusiasm to get things done		

Note. PSQI = Pittsburgh Sleep Quality Index; α = ordinal alpha coefficient; ω = ordinal omega coefficient.
[a] Values ≥ .4 represent excellent item quality (Ebel & Frisbie, 1991).

Table 2
Reliability Coefficients of the Brief Version of the Pittsburgh Sleep Quality Index (B-PSQI) Questions

B-PSQI questions	*R* item total[a]
When have you usually gone to bed at night?[b]	.752[c]
When have you usually gotten up in the morning?[b]	
How long has it usually taken you to fall asleep each night?	.509
How many hours of actual sleep did you get at night?	.641
Have you had trouble sleeping because you wake up in the middle of the night or early morning?	.588
How would you rate your sleep quality overall?	.780

Note.
[a] Item-total scale correlations, corrected for item overlap and scale reliability. [b] Rise-time and bedtime questions are used to calculate sleep-efficiency component. [c] Item-total scale correlation of sleep efficiency.

CFA was also performed by correlating residual scores between sleep efficiency and hours of sleep because of their internal communalities (Ho & Fong, 2014; Raniti et al., 2018; see Figure 2). Error covariance CFA showed goodness-of-fit improvement for the three models, indicating satisfactory fitting for the PSQI-COMP model, $\chi^2(13) = 50.418$, $p < .05$, CFI = 0.99, NFI = 0.98, TLI = 0.98, RMSEA = 0.05, SRMR = 0.06, and optimal fit for the B-PSQI model, $\chi^2(4) = 22.428$, $p < .01$, CFI = 0.99, NFI = 0.99, TLI = 0.98, RMSEA = 0.06, SRMR = 0.04 (see Table 3).

Measurement Invariance

The error covariance B-PSQI model was selected to perform measurement invariance across gender and age because it provided the best-fit model in CFA while reducing the information shared between sleep efficiency and hours of sleep. This model showed acceptable fit for men, $\chi^2(4) = 8.535$, $p > .05$, CFI = 0.99, RMSEA = 0.037, SRMR = 0.047, and women, $\chi^2(4) = 14.772$, $p > .01$, CFI = 0.99, RMSEA = 0.049, SRMR = 0.037, and also for the groups of young adults, $\chi^2(4) = 23.362$, $p < .01$, CFI = 0.98, RMSEA = 0.09, SRMR = 0.048; middle-aged adults, $\chi^2(4) = 7.200$, $p > .05$, CFI = 0.99, RMSEA = 0.02, SRMR = 0.036; and older adults, $\chi^2(4) = 8.469$, $p > .05$, CFI = 0.99, RMSEA = 0.04, SRMR = 0.047. The results for the invariance models' fit are presented in Table 4.

Invariance across gender. The configural invariance model reached satisfactory values of fit indexes (CFI = 0.99, RMSEA = .079, SRMR = 0.028), which suggested that the unifactorial structure of sleep quality applied to men and women equally. The comparison of configural and metric models showed nonsignificant $\Delta\chi^2 (p > .05)$, with low values of ΔCFI, ΔRMSEA, and ΔSRMR. These findings indicate that the fit of metric invariance did not change significantly from the configural model, and therefore items' weights were similar in both groups.

On the other hand, differences between metric and scalar models showed that $\Delta\chi^2$ was statistically significant ($p < .05$), which indicated a substantial decrease in model fit. These results suggested noninvariance of thresholds across gender, meaning that at least one item measured responses differently in men and women. To identify which items had noninvariant thresholds, we examined partial invariance. The resultant partial invariance model contained all B-PSQI items fixed except one: "Have you had trouble sleeping because you wake up in the middle of the night or early in the morning?" As Table 4 reports, the $\Delta\chi^2$ of the partial invariance model was nonsignificant ($p > .05$) compared with the metric

Table 3
One-Factor Model Goodness-of-Fit Indexes of CFA for the Original Pittsburgh Sleep Quality Index (PSQI) and the Brief Version (B-PSQI)

Models	SB-χ^2 (*df*)	CFI	NFI	TLI	RMSEA [95% CI]	SRMR
Standard CFA						
PSQI-ITEM	767.532** (119)	.879	.858	.862	.090 [.083, .096]	.100
PSQI-COMP	82.373** (14)	.973	.964	.959	.069 [.050, .089]	.073
B-PSQI	39.865** (5)	.969	.983	.937	.078 [.048, .111]	.050
Error covariance CFA						
PSQI-ITEM	709.046** (118)	.891	.870	.874	.086 [.079, .092]	.097
PSQI-COMP	50.418** (13)	.987	.978	.979	.049 [.028, .071]	.061
B-PSQI	22.428** (4)	.994	.991	.984	.060 [.024, .098]	.039

Note. CFA = confirmatory factor analysis; SB-χ^2 = Satorra–Bentler scaled chi-square; *df* = degrees of freedom; CFI = comparative fit index scaled; NFI = normed fit index scaled; TLI = Tucker–Lewis index scaled; RMSEA = root mean squared error of approximation with coefficient intervals; SRMR = standardized root mean square residual; PSQI-ITEM = CFA model of the original PSQI calculated using its 18 items; PSQI-COMP = CFA model of the original PSQI calculated using the 7 sleep components; B-PSQI = CFA model of the developed 6-item Brief Version of the PSQI.
** $p < .01$.

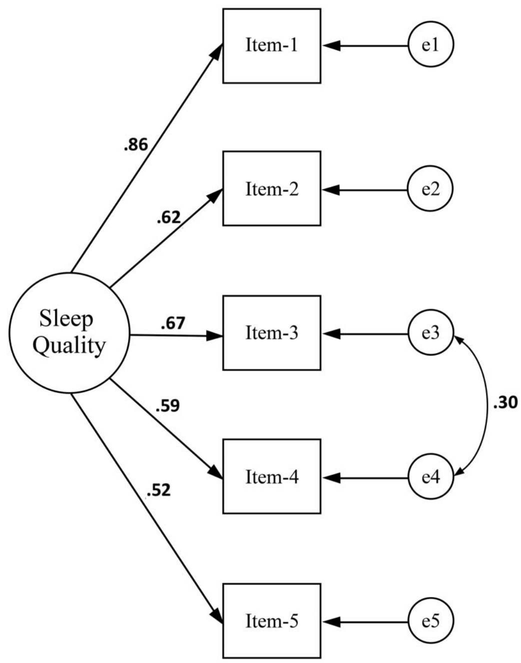

Figure 2. Standardized factor loadings of confirmatory factor analysis for the B-PSQI and residual scores correlations between sleep efficiency and hours of sleep. Item-1: sleep quality; Item-2: night awakenings; Item-3: sleep efficiency; Item-4: hours of sleep; Item-5: sleep latency.

model, and the overall fit criteria of the partial invariance model were adequate (ΔCFI = 0.007, ΔRMSEA = 0.03, ΔSRMR = 0.001), suggesting minor fit change. Because of the small number of parameters released from the model (no more than 20%), partial strong invariance across gender was considered satisfactory (Dimitrov, 2010).

Invariance across age. From a global examination, the model for configural invariance showed satisfactory goodness of fit (CFI = 0.98, RMSEA = 0.105, SRMR = 0.036). RMSEA values (above cutoff 0.08) were inconsistent with the CFI and SRMR, which may be due to data distribution and fit function rather than a problem in model fit (Lai & Green, 2016). Therefore, the structure of the B-PSQI appeared to be invariant across the three age groups. The comparison of configural and metric models showed a significant $\Delta\chi^2$ statistic ($p < .01$) and a difference in CFI values sufficiently large (ΔCFI = 0.018) to consider a major decrease in the model fit (see Table 4). Because metric invariance was not supported, we did not proceed to the scalar testing.

Convergent and Concurrent Validity

Spearman correlations between sleep self-reported measures indicated that the PSQI and B-PSQI were highly correlated (r = .895) and that high scores on these measures were significantly associated ($p < .01$) with high scores on the ISI (r = .738 for PSQI and r = .671 for B-PSQI).

As shown in Table 5, the PSQI and the B-PSQI yielded AUC values over the threshold of acceptable discrimination power (0.882 and 0.846, respectively) to identify people reporting sleep problems, indicated by ISI scores ≥ 8.

ROC analysis indicated that scores on the B-PSQI of > 5 were optimal to classify poor sleep quality, maximizing rates of sensitivity (75.82%) and specificity (76.99%). For the original PSQI version, the optimal cutoff point to classify poor sleepers was > 6, demonstrating similar rates of sensitivity (82.38%) and specificity (76.71%) compared with the B-PSQI (see Table 5). These cutoff points (B-PSQI > 5 and PSQI > 6) showed high PPV, NPV, and Youden index values associated with the minimum proportion of true positives recommended for screening purposes.

Using these cutoff values, the base rate of poor-quality sleepers was 47% (n = 286) for the PSQI (score > 6) and 44.2% (n = 269) for the B-PSQI (score > 5). As shown in Figure 3, the two measures differed in classifying 4.8% (n = 43) of poor sleepers and 5.2% (n = 32) of good sleepers, suggesting that the PSQI and the B-PSQI are statistically and clinically analogous.

Discussion

Given the relevance of the PSQI and the benefits of short questionnaires in research and clinical practice, the aim of this study was to develop the shortest PSQI version that could provide adequate validity and reliability properties in a population-based sample. The B-PSQI reduces the number of questions by 70%, going from the 18 items of the original version to 6 items. This study demonstrated that the new six-item B-PSQI is considerably shorter than the short version proposed by Famodu et al. (2018) and that it has satisfactory psychometric properties in terms of

Table 4
Measurement Invariance in the Brief Version of the Pittsburgh Sleep Quality Index (B-PSQI) Across Gender and Age

Model	SB-χ^2 (df)	CFI	RMSEA [95% CI]	SRMR	$\Delta\chi^2$(df)	ΔCFI	ΔRMSEA	ΔSRMR
Gender invariance								
Configural	23.204 (8)**	.986	.079 [.043, .117]	.028	—	—		—
Metric (λs)	25.421 (12)*	.988	.061 [.027, .093]	.030	4.364 (4)	.002	.018	.002
Scalar (λs + τs)	43.693 (21)**	.980	.060 [.034, .084]	.032	19.203 (9)*	.008	.001	.002
Partial scalar invariance	25.884 (20)	.995	.031 [.001, .062]	.031	7.231 (8)	.007	.030	.001
Age invariance								
Configural	39.083 (12)**	.978	.105 [.070, .143]	.036	—	—	—	—
Metric (λs)	68.708 (20)**	.960	.110 [.082, .138]	.046	30.199 (8)**	.018	.005	.01

Note. SB-χ^2 = Satorra–Bentler scaled chi-square; df = degrees of freedom; CFI = comparative fit index; RMSEA = root mean squared error of approximation; SRMR = standardized root mean square residual; $\Delta\chi^2(df)$ = χ^2 difference test per degree of freedom; ΔCFI = models' CFI difference; ΔRMSEA = models' RMSEA difference; ΔSRMR = models' SRMR difference.
* $p < .05$. ** $p < .01$.

Table 5
Sensitivity and Specificity of the Pittsburgh Sleep Quality Index (PSQI) and the Brief Version (B-PSQI) According to the Insomnia Severity Index (ISI; ≥ 8)

Measure	Cutoff points	Sensitivity,[a] %	Specificity,[b] %	Youden index	PPV, %	NPV, %	Base rate,[c] % (*n*)
B-PSQI AUC = .846							
	≥1	≤100	≤18.08	.173	≤49.8	≥96.4	≥88.8 (541)
	>2	97.54	32.33	.299	54.1	94.1	79.6 (485)
	>3	93.03	49.59	.426	60.2	89.7	67.5 (411)
	>4	84.43	66.03	.505	67	83.8	54.2 (330)
	>5	**75.82**	**76.99**	**.528**	**72.9**	**79.6**	**44.2 (269)**
	>6	64.75	85.21	.500	78.2	74.7	34.8 (212)
	>7	56.15	93.15	.493	87	72.2	26.6 (162)
	>8	36.89	95.34	.322	86.6	64.9	17.6 (107)
	>9	≤23.77	≥97.81	.216	≥89.9	≤61.1	≤10.8 (66)
PSQI AUC = .882							
	≥1	≤100	≤4.38	≤.044	≤46.1	100	≥97.4 (593)
	>2	100	15.62	.156	49.2	100	90.6 (552)
	>3	99.59	29.32	.289	53.5	98.9	82.3 (501)
	>4	95.49	48.77	.443	60.4	93	69.0 (420)
	>5	89.34	65.21	.545	67.8	88.2	56.7 (345)
	>6	**82.38**	**76.71**	**.591**	**74.3**	**84.2**	**47.0 (286)**
	>7	69.26	88.22	.575	82.8	77.8	34.8 (212)
	>8	58.61	93.97	.526	88.8	73.5	27.1 (165)
	>9	≤45.49	≥96.99	≤.425	≥92.5	≤68.5	≤20.0 (122)

Note. PPV = positive predictive value; NPV = negative predictive value; AUC = area under the curve. Bold values indicate optimal cutoff performance. [a] Sensitivity = true-positive rate. [b] Specificity = true-negative rate. [c] Base rate = percentage of people referring poor sleep quality based on B-PSQI and PSQI cutoffs.

internal reliability, validity, and ability to discriminate between poor and good sleepers. Moreover, the B-PSQI has simpler and more straightforward scoring than the original version, which improves the efficiency of its use.

Despite the small number of items in the B-PSQI, reliability results maintained good internal consistency (α = .79 and ω = 0.91), agreeing with PSQI validation studies targeting a nonclinical population, where internal consistency ranges from α = .67 to 0.77 (Magee, Caputi, Iverson, & Huang, 2008; Tomfohr et al., 2013). The reduction of items discarded two of the seven PSQI sleep components: daytime dysfunction and the use of sleeping medication. Similar to previous studies, the measurement properties of the PSQI improved when the medicine-use component was excluded (Mollayeva et al., 2016) or clustered together with the daytime-dysfunction component in a two-factor model (Kotronoulas, Papadopoulou, Papapetrou, & Patiraki, 2011). Daytime dysfunction and use of sleeping medication appear to have the greatest divergence from other PSQI components on conceptual and content grounds. Other sleep components, such as sleep latency or sleep disturbances, seemed to be reliably assessed with fewer items, optimizing the efficiency of the instrument. Interestingly, the largest item-total correlation for the B-PSQI was for the subjective sleep-quality component (0.78), which also shows the largest item-total correlation for the original PSQI (Buysse et al., 1989).

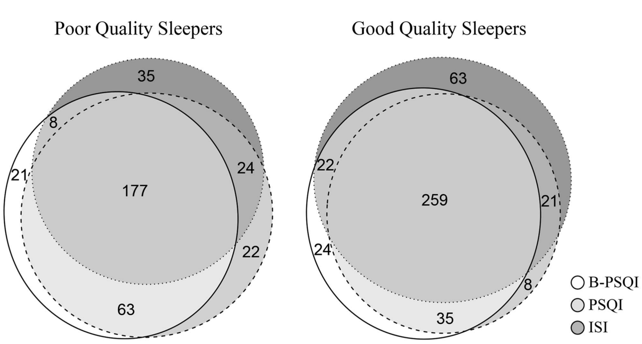

Figure 3. Venn diagram of the overlaps of good and bad sleepers (n) categorized by the Brief Version of the Pittsburgh Sleep Quality Index (B-PSQI) (score > 5), the Pittsburgh Sleep Quality Index (PSQI) (score > 6) and the Insomnia Severity Index (ISI) (score ≥ 8).

Regarding validity, CFA results showed adequate fit for the B-PSQI unidimensional structure, supporting sleep quality as a single construct, as observed in several PSQI validation studies (de la Vega et al., 2015; Manzar et al., 2018; Zhu et al., 2018). This unifactorial structure differs from the six-factor PSQI short version (Famodu et al., 2018) and from other multifactorial structure validations that have been demonstrated to yield good data representation (Mollayeva et al., 2016). The B-PSQI, with only six questions, would unlikely demonstrate adequate multifactor structure because it falls short of the recommended minimum number of three items per factor (Kline, 1998) and because of the elimination of the daytime dysfunction and medication components.

Furthermore, the results of this study indicate that the B-PSQI performs equally with men and women but only in similar age groups. Partial strong invariance was achieved across gender, suggesting that the B-PSQI provides valid mean scores in men and women and that their scores' differences reflect true differences in the sleep-quality construct. On the other hand, because age invariance was only achieved for factor structure, the comparison of B-PSQI scores is valid only among people of similar age.

These results concur with the original PSQI, where partial invariance across gender is achieved (Li, Sheehan, & Thompson, 2019), but invariance across age is only satisfactory at the configural level (Jia, Chen, Deutz, Bukkapatnam, & Woltering, 2019). The lack of metric and scalar invariance across age suggests that at least one B-PSQI item is more closely related to poor sleep in a particular age group than in another and/or that the thresholds cutoffs in the quantitative items (e.g., sleep latency) are not adapted to age-related sleep changes (Gadie et al., 2017; Hinz et al., 2017). Researchers should take into consideration these invariance results because the differences found could be due to measurement limitations (Byrne, 2008).

The B-PSQI was significantly related to PSQI and ISI scores, indicating concurrent and convergent validity, which coincides with previous PSQI validity studies (Chiu et al., 2016; Morin et al., 2011). Moreover, although the reduction of items could affect the psychometric properties of the B-PSQI, sensitivity and specificity rates were similar to the original PSQI version, with a B-PSQI cutoff of 5. This analogy with the original PSQI is also supported by the results of previous studies in which the percentage of poor sleepers among a nonclinical Spanish population (PSQI = 38.2%) was similar to the rates obtained in our study (B-PSQI = 44.2%) with similar sample characteristics (Madrid-Valero et al., 2017).

The B-PSQI seems to adequately represent sleep quality, showing slightly better adjustment for unifactorial structure than the original PSQI. The B-PSQI may provide benefits for large surveys and screening studies by improving the efficiency of assessment while preserving the ability to identify individuals with sleep problems or associated disorders and pathologies (Baglioni et al., 2016; Jike et al., 2018).

Interestingly, the B-PSQI includes mainly quantitative variables (efficiency, hours of sleep, sleep latency). Although self-reported sleep times are imprecise, these variables correspond with the assessment of sleep quality using objective measures such as actigraphy or polysomnography (Ibáñez et al., 2018; Svetnik et al., 2020). In this regard, the B-PSQI differs from other short self-reported sleep measures, such as the eight-item short-form PROMIS sleep scale and the one-item SQS, that include only graded qualitative items. The collection of sleep times allows for standardized sleep criteria and improves operational metrics (Yu et al., 2012). For example, when assessing hours of sleep, the PROMIS item "I got enough sleep" has to be answered using a Likert-type scale from *never* to *always*. However, the criterion "enough sleep" is not stipulated. By contrast, the B-PSQI item "How many hours do you sleep?" collects numeric information that allows for the standardization of a criterion for the recommended number of hours of sleep (e.g., young adults $>$ 7 hr per night). Because of these differences in response types, the B-PSQI could potentially be used to identify sleep problems based on self-reported quantitative parameters, such as advanced sleep-phase disorder.

The B-PSQI excludes items relating to daytime symptoms, such as sleepiness, which may limit a relevant part of the sleep-quality construct (Buysse, 2014; Goelema et al., 2018; Ramlee et al., 2017). However, the item of the overall perception of sleep quality may in part reflect respondents' judgments about sleep impact during the waking period, as occurs with the one-item SQS (Snyder et al., 2018).

Several limitations are worth noting. First, this study was performed among the Spanish population, which could limit B-PSQI validity in other populations. However, the use of the original PSQI has demonstrated to be invariant across both English and Spanish languages and across different cultures (Otte et al., 2013; Tomfohr et al., 2013). This suggests that the derivative brief form may be invariant as well; further validation studies are necessary to ensure this. Second, the use of the snowball technique can lead to oversampling of a particular network, as suggested by the relatively high educational attainment of the current sample. However, the online recruitment method has been used to achieve large and diverse population samples cost-effectively in other studies (Christensen et al., 2017; Kesse-Guyot et al., 2013). Furthermore, as already mentioned, the percentage of poor sleepers in this study concurs with previous studies, regardless of sampling method (Madrid-Valero et al., 2017).

In conclusion, the B-PSQI is a reliable and valid sleep-quality measure for the general population that allows easy and rapid administration. The B-PSQI may be useful as a screening tool among researchers and general practitioners because it requires little time and may maximize response rates in questionnaires, which can mitigate response-bias effects (Galesic & Bosnjak, 2009; Rolstad et al., 2011). For instance, the B-PSQI may be a useful instrument to include in national health surveys, as well as to assess sleep-related symptomatology in psychological disorders and medical problems. Future investigations in different populations should be conducted to confirm these findings.

References

Adams, R. J., Appleton, S. L., Taylor, A. W., Gill, T. K., Lang, C., McEvoy, R. D., & Antic, N. A. (2017). Sleep health of Australian adults in 2016: Results of the 2016 Sleep Health Foundation national survey. *Sleep Health, 3,* 35–42. http://dx.doi.org/10.1016/j.sleh.2016.11.005

Ainur, A. K., Sayang, M. D., Jannoo, Z., & Yap, B. W. (2017). Sample size and non-normality effects on goodness of fit measures in structural equation models. *Pertanika Journal of Science & Technology, 25,* 575–586.

Babson, K. A., Blonigen, D. M., Boden, M. T., Drescher, K. D., & Bonn-Miller, M. O. (2012). Sleep quality among U.S. military veterans with PTSD: A factor analysis and structural model of symptoms. *Journal of Traumatic Stress, 25,* 665–674. http://dx.doi.org/10.1002/jts.21757

Baglioni, C., Nanovska, S., Regen, W., Spiegelhalder, K., Feige, B., Nissen, C., . . . Riemann, D. (2016). Sleep and mental disorders: A meta-analysis of polysomnographic research. *Psychological Bulletin, 142,* 969–990. http://dx.doi.org/10.1037/bul0000053

Bastien, C. H., Vallières, A., & Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine, 2,* 297–307. http://dx.doi.org/10.1016/S1389-9457(00)00065-4

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246. http://dx.doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606. http://dx.doi.org/10.1037/0033-2909.88.3.588

Bowen, N. K., & Masa, R. D. (2015). Conducting measurement invariance tests with ordinal data: A guide for social work researchers. *Journal of the Society for Social Work and Research, 6,* 229–249. http://dx.doi.org/10.1086/681607

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21,* 230–258. http://dx.doi.org/10.1177/0049124192021002005

Buysse, D. J. (2014). Sleep health: Can we define it? Does it matter? *Sleep: Journal of Sleep and Sleep Disorders Research, 37,* 9–17. http://dx.doi.org/10.5665/sleep.3298

Buysse, D. J., Reynolds, C. F., III, Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research, 28,* 193–213. http://dx.doi.org/10.1016/0165-1781(89)90047-4

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema, 20,* 872–882.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504. http://dx.doi.org/10.1080/10705510701301834

Chiu, H. Y., Chang, L. Y., Hsieh, Y. J., & Tsai, P. S. (2016). A meta-analysis of diagnostic accuracy of three screening tools for insomnia. *Journal of Psychosomatic Research, 87,* 85–92. http://dx.doi.org/10.1016/j.jpsychores.2016.06.010

Christensen, T., Riis, A. H., Hatch, E. E., Wise, L. A., Nielsen, M. G., Rothman, K. J., . . . Mikkelsen, E. M. (2017). Costs and efficiency of online and offline recruitment methods: A web-based cohort study. *Journal of Medical Internet Research, 19,* e58. http://dx.doi.org/10.2196/jmir.6716

Corlateanu, A., Covantev, S., Botnaru, V., Sircu, V., & Nenna, R. (2017). To sleep, or not to sleep—That is the question, for polysomnography. *Breathe, 13,* 137–140. http://dx.doi.org/10.1183/20734735.007717

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98–104. http://dx.doi.org/10.1037/0021-9010.78.1.98

de la Vega, R., Tomé-Pires, C., Solé, E., Racine, M., Castarlenas, E., Jensen, M. P., & Miró, J. (2015). The Pittsburgh Sleep Quality Index: Validity and factor structure in young people. *Psychological Assessment, 27*(4), e22–e27. http://dx.doi.org/10.1037/pas0000128

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43,* 121–149. http://dx.doi.org/10.1177/0748175610373459

Ebel, R., & Frisbie, D. (1991). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice Hall.

Famodu, O. A., Barr, M. L., Holásková, I., Zhou, W., Morrell, J. S., Colby, S. E., & Olfert, M. D. (2018). Shortening of the Pittsburgh Sleep Quality Index survey using factor analysis. *Sleep Disorders, 2018,* 9643937. http://dx.doi.org/10.1155/2018/9643937

Fernandez-Mendoza, J., Rodriguez-Muñoz, A., Vela-Bueno, A., Olavarrieta-Bernardino, S., Calhoun, S. L., Bixler, E. O., & Vgontzas, A. N. (2012). The Spanish version of the Insomnia Severity Index: A confirmatory factor analysis. *Sleep Medicine, 13,* 207–210. http://dx.doi.org/10.1016/j.sleep.2011.06.019

Fieller, E. C., Hartley, H. O., & Pearson, E. S. (1957). Tests for rank correlation coefficients. I. *Biometrika, 44,* 470–481. http://dx.doi.org/10.1093/biomet/44.3-4.470

Ford, E. S., Cunningham, T. J., Giles, W. H., & Croft, J. B. (2015). Trends in insomnia and excessive daytime sleepiness among U.S. adults from 2002 to 2012. *Sleep Medicine, 16,* 372–378. http://dx.doi.org/10.1016/j.sleep.2014.12.008

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17,* 1–13.

Gadie, A., Shafto, M., Leng, Y., Cam-CAN, & Kievit, R. A. (2017). How are age-related differences in sleep quality associated with health outcomes? An epidemiological investigation in a UK cohort of 2406 adults. *British Medical Journal Open, 7,* e014920. http://dx.doi.org/10.1136/bmjopen-2016-014920

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73,* 349–360. http://dx.doi.org/10.1093/poq/nfp031

Goelema, M. S., de Bruijn, R., Overeem, S., Møst, E., Haakma, R., & Markopoulos, P. (2018). Conceptions of sleep experience: A layman perspective. *BMC Research Notes, 11,* 494. http://dx.doi.org/10.1186/s13104-018-3584-2

Hinz, A., Glaesmer, H., Brähler, E., Löffler, M., Engel, C., Enzenbach, C., . . . Sander, C. (2017). Sleep quality in the general population: Psychometric properties of the Pittsburgh Sleep Quality Index, derived from a German community sample of 9284 people. *Sleep Medicine, 30,* 57–63. http://dx.doi.org/10.1016/j.sleep.2016.03.008

Ho, R. T. H., & Fong, T. C. T. (2014). Factor structure of the Chinese version of the Pittsburgh Sleep Quality Index in breast cancer patients. *Sleep Medicine, 15,* 565–569. http://dx.doi.org/10.1016/j.sleep.2013.10.019

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Ibáñez, V., Silva, J., & Cauli, O. (2018). A survey on sleep assessment methods. *PeerJ, 6,* e4849. http://dx.doi.org/10.7717/peerj.4849

Jia, Y., Chen, S., Deutz, N. E. P., Bukkapatnam, S. T. S., & Woltering, S. (2019). Examining the structure validity of the Pittsburgh Sleep Quality Index. *Sleep and Biological Rhythms, 17,* 209–221. http://dx.doi.org/10.1007/s41105-018-00201-0

Jike, M., Itani, O., Watanabe, N., Buysse, D. J., & Kaneita, Y. (2018). Long sleep duration and health outcomes: A systematic review, meta-analysis and meta-regression. *Sleep Medicine Reviews, 39,* 25–36. http://dx.doi.org/10.1016/j.smrv.2017.06.011

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A., Rosseel, Y., Miller, P., Quick, C., . . . Mansolf, M. (2018). *semTools: Useful tools for structural equation modeling* (Version 0.5-1). Retrieved from https://CRAN.R-project.org/package=semTools

Jung, E., & Yoon, M. (2016). Comparisons of three empirical methods for partial factorial invariance: Forward, backward, and factor-ratio tests. *Structural Equation Modeling, 23,* 567–584. http://dx.doi.org/10.1080/10705511.2015.1138092

Kesse-Guyot, E., Andreeva, V., Castetbon, K., Vernay, M., Touvier, M., Méjean, C., . . . Hercberg, S. (2013). Participant profiles according to recruitment source in a large web-based prospective study: Experience from the Nutrinet-Santé Study. *Journal of Medical Internet Research, 15,* e205. http://dx.doi.org/10.2196/jmir.2488

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Kotronoulas, G. C., Papadopoulou, C. N., Papapetrou, A., & Patiraki, E. (2011). Psychometric evaluation and feasibility of the Greek Pittsburgh Sleep Quality Index (GR-PSQI) in patients with cancer receiving chemotherapy. *Supportive Care in Cancer, 19,* 1831–1840. http://dx.doi.org/10.1007/s00520-010-1025-4

Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research, 51,* 220–239. http://dx.doi.org/10.1080/00273171.2015.1134306

Landry, G. J., Best, J. R., & Liu-Ambrose, T. (2015). Measuring sleep quality in older adults: A comparison using subjective and objective methods. *Frontiers in Aging Neuroscience, 7,* 166. http://dx.doi.org/10.3389/fnagi.2015.00166

Li, L., Sheehan, C. M., & Thompson, M. S. (2019). Measurement invariance and sleep quality differences between men and women in the Pittsburgh Sleep Quality Index. *Journal of Clinical Sleep Medicine, 15,* 1769–1776. http://dx.doi.org/10.5664/jcsm.8082

Madrid-Valero, J. J., Martínez-Selva, J. M., do Couto, B. R., Sánchez-Romera, J. F., & Ordoñana, J. R. (2017). Age and gender effects on the prevalence of poor sleep quality in the adult population. *Gaceta Sanitaria, 31,* 18–22. http://dx.doi.org/10.1016/j.gaceta.2016.05.013

Magee, C. A., Caputi, P., Iverson, D. C., & Huang, X.-F. (2008). An investigation of the dimensionality of the Pittsburgh Sleep Quality Index in Australian adults. *Sleep and Biological Rhythms, 6,* 222–227. http://dx.doi.org/10.1111/j.1479-8425.2008.00371.x

Mallampalli, M. P., & Carter, C. L. (2014). Exploring sex and gender differences in sleep health: A Society for Women's Health research report. *Journal of Women's Health, 23,* 553–562. http://dx.doi.org/10.1089/jwh.2014.4816

Manzar, M. D., BaHammam, A. S., Hameed, U. A., Spence, D. W., Pandi-Perumal, S. R., Moscovitch, A., & Streiner, D. L. (2018). Dimensionality of the Pittsburgh Sleep Quality Index: A systematic review. *Health and Quality of Life Outcomes, 16,* 89. http://dx.doi.org/10.1186/s12955-018-0915-x

Mollayeva, T., Thurairajah, P., Burton, K., Mollayeva, S., Shapiro, C. M., & Colantonio, A. (2016). The Pittsburgh Sleep Quality Index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. *Sleep Medicine Reviews, 25,* 52–73. http://dx.doi.org/10.1016/j.smrv.2015.01.009

Morin, C. M., Belleville, G., Bélanger, L., & Ivers, H. (2011). The Insomnia Severity Index: Psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep: Journal of Sleep and Sleep Disorders Research, 34,* 601–608. http://dx.doi.org/10.1093/sleep/34.5.601

Otte, J. L., Rand, K. L., Carpenter, J. S., Russell, K. M., & Champion, V. L. (2013). Factor analysis of the Pittsburgh Sleep Quality Index in breast cancer survivors. *Journal of Pain and Symptom Management, 45,* 620–627. http://dx.doi.org/10.1016/j.jpainsymman.2012.03.008

Pintea, S., & Moldovan, R. (2009). The receiver-operating characteristic (ROC) analysis: Fundamentals and applications in clinical psychology. *Journal of Cognitive and Behavioral Psychotherapies, 9,* 49–66.

Ramlee, F., Sanborn, A. N., & Tang, N. K. Y. (2017). What sways people's judgment of sleep quality? A quantitative choice-making study with good and poor sleepers. *Sleep: Journal of Sleep and Sleep Disorders Research, 40,* zsx091. http://dx.doi.org/10.1093/sleep/zsx091

Raniti, M. B., Waloszek, J. M., Schwartz, O., Allen, N. B., & Trinder, J. (2018). Factor structure and psychometric properties of the Pittsburgh Sleep Quality Index in community-based adolescents. *Sleep: Journal of Sleep and Sleep Disorders Research, 41,* zsy066. http://dx.doi.org/10.1093/sleep/zsy066

Rener-Sitar, K., John, M. T., Bandyopadhyay, D., Howell, M. J., & Schiffman, E. L. (2014). Exploration of dimensionality and psychometric properties of the Pittsburgh Sleep Quality Index in cases with temporomandibular disorders. *Health and Quality of Life Outcomes, 12,* 10. http://dx.doi.org/10.1186/1477-7525-12-10

Revelle, W. (2018). *psych: Procedures for psychological, psychometric, and personality research* (Version 1.8.10). Retrieved from https://CRAN.R-project.org/package=psych

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health, 14,* 1101–1108. http://dx.doi.org/10.1016/j.jval.2011.06.003

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48,* 1–36. http://dx.doi.org/10.18637/jss.v048.i02

Royuela, A., & Macías Fernández, J. (1997). Propiedades clinimetricas de la versión castellana del cuestionario de Pittsburgh [Clinimetric properties of the Spanish version of the Pittsburgh Questionnaire]. *Vigilia-Sueño, 9,* 81–94.

Rutkowski, L., & Svetina, D. (2017). Measurement invariance in international surveys: Categorical indicators and fit measure performance. *Applied Measurement in Education, 30,* 39–51. http://dx.doi.org/10.1080/08957347.2016.1243540

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507–514. http://dx.doi.org/10.1007/BF02296192

Snyder, E., Cai, B., DeMuro, C., Morrison, M. F., & Ball, W. (2018). A new single-item Sleep Quality Scale: Results of psychometric evaluation in patients with chronic primary insomnia and depression. *Journal of Clinical Sleep Medicine, 14,* 1849–1857. http://dx.doi.org/10.5664/jcsm.7478

Stranges, S., Tigbe, W., Gómez-Olivé, F. X., Thorogood, M., & Kandala, N.-B. (2012). Sleep problems: An emerging global epidemic? Findings from the INDEPTH WHO-SAGE study among more than 40,000 older adults from 8 countries across Africa and Asia. *Sleep: Journal of Sleep and Sleep Disorders Research, 35,* 1173–1181. http://dx.doi.org/10.5665/sleep.2012

Svetnik, V., Snyder, E. S., Tao, P., Roth, T., Lines, C., & Herring, W. J. (2020). How well can a large number of polysomnography sleep measures predict subjective sleep quality in insomnia patients? *Sleep Medicine, 67,* 137–146. http://dx.doi.org/10.1016/j.sleep.2019.08.020

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science, 240,* 1285–1293. http://dx.doi.org/10.1126/science.3287615

Tomfohr, L. M., Schweizer, C. A., Dimsdale, J. E., & Loredo, J. S. (2013). Psychometric characteristics of the Pittsburgh Sleep Quality Index in English speaking Non-Hispanic whites and English and Spanish speaking Hispanics of Mexican descent. *Journal of Clinical Sleep Medicine, 9,* 61–66. http://dx.doi.org/10.5664/jcsm.2342

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology, 7,* 769. http://dx.doi.org/10.3389/fpsyg.2016.00769

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10. http://dx.doi.org/10.1007/BF02291170

Wejnert, C., & Heckathorn, D. D. (2008). Web-based network sampling: Efficiency and efficacy of respondent-driven sampling for online research. *Sociological Methods & Research, 37,* 105–134. http://dx.doi.org/10.1177/0049124108318333

Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. M. (2011). On creating and using short forms of scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, & R. E. Lucas (Eds.), *Secondary data analysis: An introduction for psychologists* (pp. 39–61). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/12350-003

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling, 17,* 392–423. http://dx.doi.org/10.1080/10705511.2010.489003

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3,* 32–35. http://dx.doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Yu, L., Buysse, D. J., Germain, A., Moul, D. E., Stover, A., Dodds, N. E., . . . Pilkonis, P. A. (2012). Development of short forms from the PROMIS Sleep Disturbance and Sleep-Related Impairment item banks. *Behavioral Sleep Medicine, 10,* 6–24. http://dx.doi.org/10.1080/15402002.2012.636266

Zhu, B., Xie, M., Park, C. G., & Kapella, M. C. (2018). Adaptation of the Pittsburgh Sleep Quality Index in Chinese adults with Type 2 diabetes. *Journal of the Chinese Medical Association, 81,* 242–247. http://dx.doi.org/10.1016/j.jcma.2017.06.021

Zomers, M. L., Hulsegge, G., van Oostrom, S. H., Proper, K. I., Verschuren, W. M., & Picavet, H. S. J. (2017). Characterizing adult sleep behavior over 20 years—The population-based Doetinchem Cohort Study. *Sleep: Journal of Sleep and Sleep Disorders Research, 40,* zsx085. http://dx.doi.org/10.1093/sleep/zsx085

Received October 11, 2019
Revision received July 31, 2020
Accepted August 17, 2020 ■