neurozone®

The Development & Validation of the Neurozone® Sleep Index

Introduction

The measurement of sleep disturbance is both a vast and variegated topic in the scientific literature. Most of the recent papers specify the measurement of sleep disturbance by particular comorbid conditions (e.g., in **psoriasis**, **spinal cord injury**, and **mild cognitive impairment**), and/or by specific demographic or professional profile (e.g, in **the elderly**, **nurses**, and **military personnel**). While these cohort-specific measurements are valuable, recent research appears to be wanting in the measurement of sleep disturbances in the general population. Moreover, many measures developed in high-income countries show limited applicability to populations in low- to middle-income countries (such as **Sri Lanka** and **Arabic nations**). There is thus also a need for a measurement of sleep disturbance in the general population that can apply to both high- and low- to middle-income countries. Finally, as detailed below, many of the existing measures tend to carry limitations in the nature and extent of the data they collect, making it necessary to refine the measurement of sleep disturbance so as to increase the utility of the data collected.

Scale Development

As mentioned above, although there are various existing measures of sleep quality, there are several significant psychometric limitations associated with a large number of these questionnaires. Firstly, one major limitation relates to the Likert-type nature of these questionnaires *coupled* with a low number of response options (typically fewer than 5). This results in ordinal data that in most instances cannot be treated as continuous (due to the limited number of response options). This raises questions around the validity and the generalizability of the scale development process of these questionnaires (e.g., using exploratory factor analysis and confirmatory factor analysis), as there is often limited or no disclosure on whether corrections were used such as, for example, alternative correlation matrices and different estimators and extraction methods. Secondly, and relatedly, data of this type presents a major obstacle in terms of being able to (reliably) include it in predictive analysis with, for example, other psychiatric measures. This limits an accurate and reliable understanding of how common psychiatric symptoms relate to one another in a predictive manner. Thirdly, existing questionnaires do not take differing sleep onset latency and total sleep time during the week and over the weekend into consideration. This likely leads to unreliable reporting/estimates. Asking about these metrics separately and averaging them during scoring, for example, could be a more reliable estimation of true sleep onset latency and total sleep time. Fourthly, most existing questionnaires have limited or unknown utility in low- to middle-income settings.

Given the above, we set out to develop a brief, valid, and reliable measure of sleep disturbance that can be used in both high- and low- to middle-income countries and that will yield continuous data that can reliably be used in predictive analysis. With these aims in mind, we conducted a thorough and up-to-date review of current empirical evidence describing the critical components that make up sleep disturbance. Upon conclusion of the review, a list of relevant constructs were identified and operationalized by an expert panel of psychologists, and finally populated with their respective items. Data collection on the Neurozone® Sleep Index commenced in 2023. The original list of items consisted of 18 items.

The Sample

The total sample consisted of 486 participants. The average age of the sample was 52 years (range = 21–81). The gender distribution of the total sample consisted of 69% men and 31% women.

Methods

Exploratory & Confirmatory Factor Analysis

Due to the relatively large sample size (486), the decision was made to randomly split the dataset in half and to perform exploratory factor analysis (EFA) on the first subset and confirmatory factor analysis (CFA) on the second subset. This is a robust approach in terms of determining and confirming a reliable and stable factor structure. By conducting EFA on one half of the dataset, one is able to explore the underlying factor structure without preconceived predictions, allowing the factors to emerge from the data mathematically. Subsequently, performing CFA on the other half of the dataset, by specifying the factor structure from EFA, enables one to statistically confirm whether the emergent factor structure replicates well in an independent dataset (i.e. the second, CFA subset). This approach also ensures homogeneity in terms of sample characteristics across the two datasets, since they were collected from the same population; however, the data points are, importantly, independent of one another and can therefore be used in separate, comparative analysis.

Data Preprocessing & Analysis

All data management and analyses were done using R Statistical Computing Software Version 2023.09.0+463. The distribution properties of 5 of the 18 original items showed limited variance across responses, rendering these items a poor representation of disrupted sleep in this sample. Given that inclusion of these items would have resulted in skewed results, we decided to exclude them from analyses. In terms of running EFA and CFA, we used the ordinary least squares (OLS) estimator with the ‘minres’ method for extraction to accommodate univariate non-normality in the data for EFA. For CFA, we used diagonally weighted least squares (DWLS) as an estimator. The DWLS is a robust estimator typically used in CFA in the context of non-normality and other potential assumption violations.

Results

Exploratory Factor Analysis

Examining the scree plot along with its associated eigenvalues revealed that a three-factor solution appears to be the most appropriate fit for the data. Although the third factor had an eigenvalue below 1.00, we examined other metrics in order to ascertain the degree to which the third factor contributed to the fit and robustness of the model. For example, we determined that the third factor increased the amount of variance explained by 14%, which significantly improved the performance of the overall model. The loadings on the third factor were also clear and large, further supporting the data-driven decision to retain the third factor.

We also examined the factor loading for each item in relation to its respective factor, which was very large in each instance (range across factors = 0.615–0.873). Based on these results we retained 10 items (generally, a retention cut-off of 0.500 is considered very robust). The cumulative variance explained was 64%. The KMO statistic as a measure of sampling adequacy was 0.810, which is considered very good. This result indicates that factor analysis was a suitable extraction method for this data.

Three-Factor Solution

The three-factor solution of the Neurozone® Sleep Index (NSI) consisted of 10 items assessing various critical components of sleep disturbance. Interestingly, despite iteratively testing different factor structures and factor-loading cut-offs before deciding on the best-fitting model, the sleep duration questions did not qualify for the final list of items.

There are several potential reasons for this: Firstly, good sleep duration does not equate to good sleep quality – many individuals reportedly get an adequate number of hours of sleep, but may experience fragmented sleep/frequent awakenings, leading to non-restorative sleep. Secondly, given that the sample in this study consisted of relatively older adults, who likely have a naturally shorter sleep duration, the results could have been skewed. Nevertheless, latest sleep science shows that **sleep quality appears to be more important than quantity** when it comes to daytime functioning. Two important determinants of sleep quality relate to sleep onset latency and sleep fragmentation, indicating that these two factors could also be important in relation to daytime functioning. Indeed, our inter-item correlational analyses have shown several strong and significant relationships between the items measuring night-time and daytime symptoms of sleep disturbance. Finally, given the aim of creating a brief, reliable, and valid questionnaire coupled with the excellent fit statistics obtained from the final model, we retained 10 items across three factors:

Factor	Description	Range of Loadings
Sleep Onset	This factor measures the amount of time it takes a person to fall asleep both in the week and over the weekend (in minutes).	0.818-0.857
Sleep Continuity	This factor assesses the lack of sleep continuity in terms of the amount of time spent awake after initially falling asleep, as well as the frequency of early-morning awakenings.	0.652-0.873
Daytime Functioning	This factor measures the degree to which a person experiences disruptive daytime symptoms in terms of fatigue, sleepiness, disrupted mood, and decreased cognitive functioning.	0.615-0.823

Reliability Analysis

We evaluated the internal consistency of the three-factor solution by computing Cronbach's alpha. We did this to evaluate the extent to which the items making up the factor actually measure the same underlying concept (e.g., sleep disturbance). Higher Cronbach's α values (which typically range from 0 to 1) indicate greater internal consistency, suggesting that the items are more reliably measuring the same underlying construct. The ideal range for Cronbach's α values, however, is between 0.70 and 0.90.

Results show the following:

- *Sleep Onset:* Cronbach's α = 0.860
- *Sleep Continuity:* Cronbach's α = 0.770
- *Daytime Functioning:* Cronbach's α = 0.890

Together, these results indicate that the Neurozone® Sleep Index has very good reliability. Other metrics from reliability analysis support this: the corrected-item total correlations for all items on their respective factors were all large and well-above the cut-off of 0.30. Furthermore, results also show that Cronbach's α did not increase in the event of any of the items being removed. This indicates that all the items included contribute to the reliability of their respective factors, while also eliminating possible concerns about scale redundancy.

Confirmatory Factor Analysis

We conducted confirmatory factor analysis on the three-factor solution derived from EFA. We computed multiple fit indices to assess both absolute and incremental goodness-of-fit. More specifically, we report on the chi-square statistic and its associated significance level, as well as the chi-square statistic divided by the degrees of freedom (X^2/df). The latter was included in order to circumvent the potential confounding effects of sample size and multivariate non-normality on the chi-square results. Other absolute indices include the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Incremental fit indices include the comparative fit index (CFI) and the Tucker-Lewis Index (TLI). The a priori cut-offs for the various fit indices include $p < 0.05$ for the chi-square statistic, a value < 5 for the X^2/df statistic, < 0.06 for the RMSEA, < 0.08 for the SRMR, and > 0.95 for the CFI and TLI.

CFA Results

Results showed that all items loaded significantly, strongly, and clearly on their respective factors with a range of 0.51 to 0.91. We calculated the participant-to-parameter ratio (PPR) as a measure of the complexity of the model relative to the amount of data available. A PRR of at least 5 to 10 is generally regarded as ideal. Results show that our PPR is 10.570, which is considered good and is indicative of adequate sample size, stable parameter estimates, and good model generalizability.

We also calculated several fit indices in order to evaluate how well the model structure fits the data. Firstly, with regard to absolute fit indices, results show that the chi-square statistic

was not statistically significant (X^2 = 34.602, p = 0.369), which indicates that the three-factor solution is a good fit for the data. This is supported by the X^2/df statistic, which yielded a value of 1.064 (cut off: < 5). Secondly, the RMSEA, as another absolute fit index, shows that the model is a close fit for the data (0.053, CI lower = 0.034, CI upper = 0.072, p = 0.366; cut-off < 0.06), while the SRMR provides additional evidence for the close fit of the three-factor solution (0.061; cut-off: < 0.080). In addition, the incremental fit statistics also indicate that the model is a good fit for the data (robust CFI = 0.998; robust TLI = 0.997).

Validity Testing

In addition to having good content validity based on the scale development methodology, we also set out to demonstrate other, objective forms of validity: concurrent validity and convergent validity.

Concurrent Validity

Concurrent validity is a form of criterion validity that measures how well a new test compares to a well-established test. In this case, to test for concurrent validity, we needed to demonstrate that there is a statistically significant relationship between the Neurozone® Sleep Index and an existing sleep disturbance measure. To this end, we assessed whether there was a significant, positive correlation between the 10-item Neurozone® Sleep Index and the existing, well-validated 8-item Athens Insomnia Scale. For both scales, a higher score is indicative of more disrupted sleep. Results showed a significant, large positive correlation (r = 0.780) between these two measures. Therefore, concurrent validity was confirmed in this study.

Convergent Validity

Convergent validity, which is a form of construct validity, refers to the extent to which two measures that theoretically should be related are, in fact, statistically related. For example, in the case of sleep disturbance, we would expect there to be a significant positive relationship with mental health-related outcomes (e.g., depression, anxiety, and burnout) and a significant *inverse* relationship with psychological resilience. Indeed, results showed significant, positive and large correlations with depression (r = 0.660), anxiety (r = 0.590), and burnout (r = 0.540). In addition, results showed a medium-to-large correlation with psychological resilience (r = 0.440). Together, these results provide convincing support for the strong convergent validity of the Neurozone® Sleep Index.

Clinical Cut-Off for Sleep Disturbance

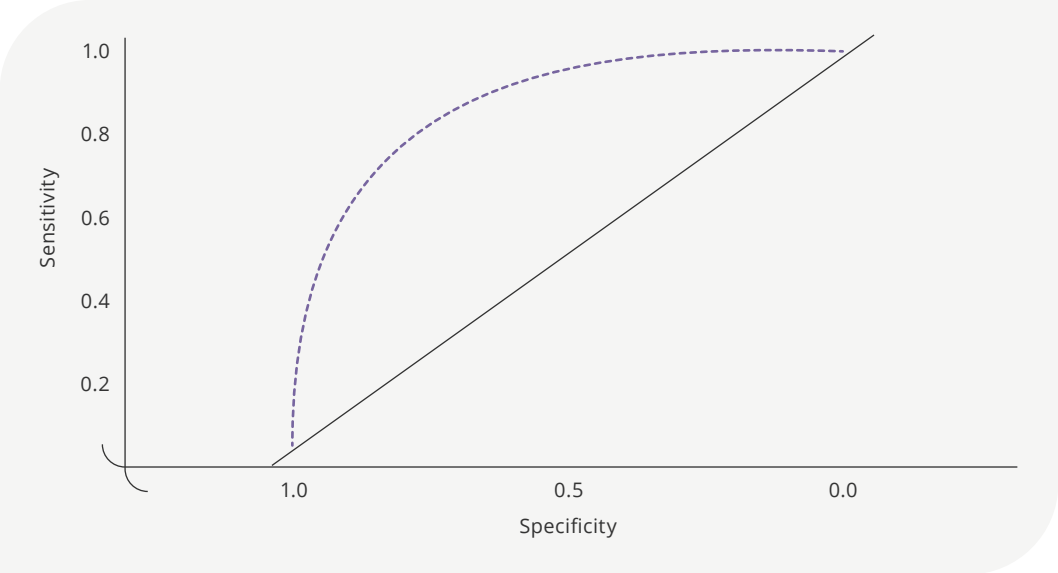
In addition to developing a new brief, reliable, and valid measure of sleep disturbance, we also set out to determine a clinical cut-off for classifying individuals as having (probable) clinically significant sleep disturbance, in order to differentiate them from those individuals who do not meet the minimum criterion. To establish this, we used the well-validated method of **Receiver Operating Characteristic (ROC) curve analysis** along with the concept of sensitivity and specificity to determine the probable clinical cut-off for sleep disturbance.

More specifically, the ROC curve is a graphical plot that illustrates the variability in diagnostic accuracy of a binary classifier system (clinical sleep disturbance: yes/no) as its discrimination threshold is adjusted. We used several key metrics in assisting us to determine the best threshold and to assess the overall performance of our model. These metrics include the area under the curve (AUC), sensitivity, specificity, and the Youden Index criterion to assist us in determining the optimal threshold. We used the Athens Insomnia Scale in the analysis as the criterion measure for determining the clinical cut-off.

ROC Analysis Results

The AUC was found to be 0.871, indicating that our model has a very high level of accuracy in distinguishing between individuals with and without clinically significant sleep disturbance:

ROC Curve for Clinical Cut-Off for Disturbed Sleep



Furthermore, results indicated that the optimal threshold for classifying clinically significant sleep disturbance is 12.963 (rounded to 13). This cut-off was determined using the Youden Index criterion coupled with manual inspection of the coordinates of the ROC curve (see above). The Youden Index, a common method for selecting the optimal cut-off point, maximizes the sum of sensitivity and specificity, ensuring the best balance between correctly identifying individuals with and without the condition. For our purposes (correctly classifying individuals with clinically significant sleep disturbances), we chose an approach where sensitivity is prioritised. Therefore, at this threshold, the sensitivity was 0.865, meaning that 86.5% of individuals with insomnia were correctly identified by the model. The specificity was 0.722, indicating that 72.2% of individuals without insomnia were correctly identified.

Taken together, these results provide strong evidence for the robustness of our diagnostic model, particularly with regard to a very high AUC and an excellent level of sensitivity. Moreover, these results indicate that our model has very good accuracy in distinguishing between individuals with and without clinically significant sleep disturbances. Importantly, the optimal threshold was identified as 13, which provides a balanced approach in terms of accurately determining a reliable clinically significant cut-off score for sleep disturbance.

Summary & Conclusion

We employed exploratory factor analysis and confirmatory factor analysis to develop a brief, valid, and reliable measure of sleep disturbance. All absolute and incremental fit indices confirmed a very good, close fit of the model to the data. In addition, we demonstrated very strong content, concurrent, and convergent validity. Importantly, our new measure can be used in both high- and low- to middle-income settings and yields continuous data suitable for predictive analyses. Finally, we determined a probable clinical cut-off for sleep disturbance that can be used to classify individuals as likely having clinically significant sleep disturbances or not. Taken together, these results provide convincing evidence for the reliability, stability, validity, and generalizability of the three-factor solution of sleep disturbance as measured by the brief 10-item Neurozone® Sleep Index.

neurozone®

www.neurozone.com