Nursing Open

WILEY

NursingOpen Open Access

EMPIRICAL RESEARCH QUANTITATIVE OPEN ACCESS

Developing a Sleepiness Scale for Nurses and Physicians: A Methodological Study

Ali Kaplan[1] | Özlem Kaplan[2] | Cevriye Özdemir[1] | Ferhan Elmalı[3] | Emre Bülbül[4]

[1]Department of Medical Services and Techniques, İncesu Ayşe and Saffet Arslan Health Services Vocational School, University of Kayseri, Kayseri, Türkiye | [2]Department Nursing, Health Science Faculty, University of Erciyes, Kayseri, Türkiye | [3]Department of Biostatistics, Faculty of Medicine, University of İzmir Kâtip Çelebi, İzmir, Türkiye | [4]Department of Emergency Medicine, Faculty of Medicine, University of Erciyes, Kayseri, Türkiye

Correspondence: Ali Kaplan (alikaplan@kayseri.edu.tr)

Received: 7 March 2024 | **Revised:** 14 February 2025 | **Accepted:** 12 August 2025

Funding: The authors received no specific funding for this work.

Keywords: nurse | physician | scale development | sleepiness

ABSTRACT

Aim: The sleepiness of healthcare workers can negatively affect various aspects of healthcare, including patient and staff safety. Therefore, the aim of this study was to develop a valid and reliable scale to measure sleepiness in nurses and physicians working night shifts.

Design: A methodological study design was used in this study.

Methods: The scale items were developed through literature review and qualitative data analysis. The initial items of the instrument were developed based on the identified core concepts and presented to 21 experts for content validity assessment. In order to test the validity and reliability of the scale, a group of 255 nurses and 230 physicians working in a university hospital in a province of Turkey were included in the study between November 2023 and January 2024 by face-to-face survey method. Lawshe's content validity ratio, Kaiser–Mayer–Olkin coefficient, Bartlett's test, exploratory factor analysis, principal component analysis, Varimax factor rotation method, confirmatory factor analysis, Cronbach's alpha internal consistency coefficient, Spearman correlation coefficient and Wilcoxon Signed Ranks tests were used to analyse the data.

Results: The sleepiness scale is a 19-item instrument with three factors: 'cognitive and physical effects of sleepiness on the process of working', 'difficulty keeping awake' and 'sleepiness coping strategies'. The Kaiser–Meyer–Olkin measures indicated that the data were sufficient for exploratory factor analysis. Bartlett's test indicated patterned relationships among the items ($\chi^2 = 5049.779$; $p < 0.001$). Three factors explained 58.927% of the variance. Internal consistency was demonstrated by Cronbach's alphas of 0.921 Factor 1, 0.766 Factor 2, 0.797 Factor 3 and 0.923 total scale. According to the confirmatory factor analysis fit index results of the scale, RMSEA value was 0.062 and χ^2/df value was 2.843. The 4-week test–retest reliability of the sleepiness scale ($n = 51$) was also satisfactory (ICC = 0.958). The Sleepiness Scale is a valid and reliable measurement tool that can be used to determine the severity of sleepiness in nurses and physicians working in the night shift.

Patient or Public Contribution: No patient or public contribution.

1 | Introduction

Sleep is an essential requirement for health and well-being and has come to be recognised as one of the leading issues in public health (Matricciani et al. 2017). Sleep, which has a rhythmic structure, can be disrupted for various reasons, leading to sleep troubles. One of these is defined as sleepiness, characterised by a tendency to fall asleep. Sleepiness is a normal physiological

condition that many individuals may experience throughout a 24-h period (Martin et al. 2023). The occurrence of sleepiness at inappropriate times, its increased frequency and reaching a level that disrupts daily functioning are considered excessive sleepiness (Smart et al. 2020).

It has been reported that excessive sleepiness is typically associated with adverse social and economic consequences, such as motor vehicle accidents, near-miss incidents, decreased work productivity and reduced quality of life (Barnes and Watson 2019; Leger and Stepnowsky 2020). It also plays a prominent role in acute cognitive performance and long-term cognitive health, including impaired attention, memory and concentration, poor performance and increased risk of dementia (Lee et al. 2022; Low et al. 2019; Ma et al. 2020; Xu et al. 2020). According to the International Classification of Sleep Disorders, it has been reported that 'sleepiness' is an important factor in many sleep disorders and sleepiness is seen in 10% of shift workers (Reid and Abbott 2015).

Many individuals work shifts in their communities due to reasons related to public safety, healthcare services, or economic factors (Gusman et al. 2023). Shift work in the healthcare sector often involves rotating schedules (8-, 16-, 24-h shifts) to provide patient care around the clock. Due to the impact of shift work on sleep rhythms, sleepiness during work has negative effects on the health and well-being of health workers (Kecklund and Axelsson 2016; Nena et al. 2018). This situation adversely affects clinical practices conducted by healthcare personnel, such as treatment, care and medication administration, thereby posing risks to both employee and patient safety (Härmä et al. 2019; Smart et al. 2020). In this context, sleepiness is recognised as a significant health and workplace issue for both nurses and physicians.

It is important to evaluate sleepiness, which can cause many negative consequences on nurses and physicians working in shifts (Härmä et al. 2019). Defining this situation will guide the interventions to be made and to achieve this goal, both subjective and objective evaluations should be made. Scales are the most well-known subjective measurement tools (South et al. 2022), and in the current literature review on sleepiness, it is seen that Cleveland Adolescent Sleepiness Questionnaire, Epworth Sleepiness Scale, Stanford Sleepiness Scale and Karolinska Sleepiness Scale are available (Geiger Brown et al. 2014; Johns 1992; Maclean et al. 1992; Spilsbury et al. 2007). It is seen that the scales available in the literature are mostly aimed at measuring the daytime sleepiness levels of individuals and are not specific to nurses and physicians. When the literature specific to nurses was examined, only the 'Bergen Shift Work Sleep Questionnaire' was found (Flo et al. 2012). However, this scale does not measure the severity of sleepiness during the shift and only allows comparison of the sleepiness levels of nurses between three shifts. Therefore, it is thought that the existing scales are insufficient to determine the severity of sleepiness of nurses and physicians during night shifts. In this context, in order to overcome this deficiency in the literature, it was aimed to develop a valid and reliable instrument to severity of sleepiness levels of nurses and physicians during night shifts.

2 | Methods

2.1 | Study Design and Procedure

In this study, a methodological approach was adopted and conducted in two main stages. In the first, the process of developing the Sleepiness Scale was comprehensively addressed: the focus was on the design of the scale and item selection, in order to establish a solid foundation for the scale. In the second stage, the validity and reliability of the scale which was developed were tested. The diagram showing the development process of the scale is given in Figure 1. The STROBE checklist was used to report the research context, study design, methods and results of the study.

2.2 | Conceptual Framework and Item Generation

The conceptual framework of the study is based on the information obtained from both the literature review and the results of qualitative data analysis. A basic framework was created to make sense of the sleepiness experience of nurses and physicians working night shifts. In this context, 'sleepiness and fatigue', 'cognitive and psychomotor performance', 'physiological and psychological effects', 'occupational and clinical outcomes' and 'coping strategies and preventive practices' were identified as the main concepts related to sleepiness. Then, existing studies that could contribute to the scale items within the framework of the basic concepts determined in the literature review on sleep and sleepiness were examined (Geiger Brown et al. 2014; Johns 1992; Maclean et al. 1992; Nena et al. 2018; Spilsbury et al. 2007). However, instead of creating an item pool based only on existing studies, a qualitative data collection process was conducted to directly understand the experiences of nurses and physicians working night shifts. In this context, individual interviews were conducted with five experienced nurses working night shifts, two physicians with academic studies on sleep and three nurse academics. During the interviews, the participants were asked 'What are the effects of sleepiness during working hours?' and 'What are the practices to prevent sleepiness during working hours?'. The data obtained was analysed using inductive content analysis (Kyngäs 2020). The analysis identified themes emerging from participants' expressions, thereby strengthening the conceptual framework of the study. Based on the conceptual framework of the study, 25 preliminary items were created using a 5-point Likert scale (Strongly Disagree, Disagree, Indecisive, Agree, Strongly Agree).

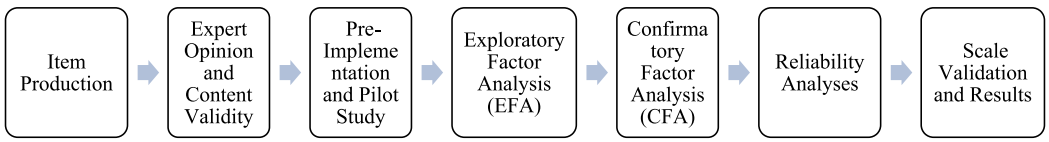


FIGURE 1 | Scale development process.

2.3 | Content Validity

The Lawshe technique was used to assess the content validity of the study. This method is a technique in which experts evaluate the relevance of each item of the measurement tool to the relevant subject or field (Romero Jeldres et al. 2023). Two rounds of content validity assessment were conducted to determine whether each of the preliminary items represented the ability to measure the sleepiness levels of both physicians and nurses. A total of 21 experts were consulted, including six academic nurses, five academic physicians, seven practicing nurse specialists and three practicing physicians, to provide their opinions on the content validity of the items. Researchers created a form consisting of expressions 'necessary and should remain in the item pool', 'useful but should be revised' and 'unnecessary' for the evaluation of each item. Additionally, a column was added for experts who marked 'useful but should be revised', where explanations and suggestions could be provided. The form created for content validity was sent to experts via email. Each item for which revision was suggested based on feedback from experts was evaluated, and necessary revisions were made. Three items were removed because they were perceived to be similar to others. For example, the item 'I make a lot of efforts not to fall asleep towards the end of my shift' was excluded from the study due to feedback indicating its similarity to 'I feel awake during the last hours of my shift'. The final version of the revised form consisting of 22 items was sent to the experts again via email, and they were asked to reassess. After receiving the final expert opinion, the content validity ratio (CVR) for each item and the Content Validity Index for the overall scale were calculated (min. CVR: 0.71, max. CVR: 1.0). CVR is a quantitative measure used to determine the content validity of a scale. It assesses how essential each item is to the scale based on expert opinion. According to the Lawshe technique, the CVR for the scale items for 21 experts should be at least 0.42 at a significance level of $\alpha = 0.05$ (Yesilyurt and Capraz 2018). Accordingly, it was determined that all items had Content Validity Index values greater than the minimum required value. Thus, it was concluded that the scale provided content validity. All items in the draft scale form consisting of 22 items, created based on expert opinions, were presented to three language experts for evaluation in terms of language and meaning. Necessary corrections were made based on the suggestions of the experts.

2.4 | Participants

To collect data, a face-to-face questionnaire was administered to nurses and physicians working at a University Hospital (tertiary care) in a province in Türkiye. The nurses and physicians working in this hospital formed the participants of the study, and a stratified sampling method was used specifically for this population. In scale validity and reliability studies, it is generally recommended to have at least 10 times the number of items in the scale ($22 \times 10 = 220$) as participants and to reach at least 10% of the accessible population size (which was at least 154 for this study) (Pallant 2020). Equal numbers of participants were planned to be recruited from nurses and physicians, with a minimum of 220 nurses and 220 physicians included in the study. To increase the generalisability of the study findings, a Systematic Random Sampling Method was also employed for strata. Efforts were made to reach similar proportions of nurses and physicians from relevant departments, such as the emergency department, intensive care unit and the clinics, according to the population of each department. During the data collection period, 255 nurses and 230 physicians were included in the study, exceeding the predetermined sample sizes. The inclusion criteria for participants were as follows: (1) actively working at the hospital where the study was conducted during the study period, (2) having at least 6 months of work experience, (3) having worked night shifts at least once in the past month, (4) who declares that he has no sleep problems, (5) not using sleep medication and (6) volunteering to participate.

2.5 | Data Collection

The data for the study was collected at a university hospital in one of the largest cities in the Central Anatolia region, between November 2023 and January 2024. An 8-item Descriptive Characteristics Information Form, containing participants' characteristics and the draft items of the developed 22-item Sleepiness Scale, was used for data collection. The data was collected face-to-face by four researchers responsible for data collection. The researchers visited the clinics and interviewed nurses and physicians during the night shift. The night shift starts at 04:00 PM in the relevant institution. The researchers visited the clinics between 08:00 and 10:00 PM and gave information about the study to the individuals participating in the study and the participants who agreed to participate in the study were asked to sign the written informed consent form. The data collection form was explained in detail to the individuals participating in the study and they were asked to complete the data collection forms, which took 5 to 10 min. After the visits to the clinics planned on the relevant day were completed, the forms were collected from the individuals in a sealed envelope by the researchers by returning to the first clinic. Test–retest analysis was performed for the reliability analysis of the draft scale (Streiner et al. 2016). For this purpose, 51 participants, 28 nurses and 23 physicians, randomly selected, were asked to fill out the draft scale again approximately 1 month later.

2.6 | Data Analysis

Statistical data analysis was performed using the IBM SPSS Statistics 26.0 package program (IBM Corp., Armonk, New York, USA). Descriptive statistics were given as frequency (n), percentage (%), mean ± standard deviation, minimum value (min) and maximum value (max). The normal distribution of the data of numerical variables was evaluated by the Shapiro–Wilk normality test and Q-Q graphs. CVRs were calculated for content validity. Construct validity, criterion-related validity, internal consistency reliability, test–retest and item analysis methods were used for the scale. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were conducted on the data obtained for construct validity. In order for explanatory factor analysis to be applied to a data group, the data must be suitable for factor analysis and the sample must be sufficient (Özdamar 2013). First, Bartlett's sphericity test for factorability and Kaiser–Meyer–Olkin (KMO) test

results were examined for the adequacy of the number of units in the sample. The reason for conducting EFA is to test the theory about the nature of the process and to make an operational definition of the basis of the process using observed variables (Tabachnick and Fidell 2014). A structure consisting of three factors was obtained through EFA, principal components approach and varimax rotation method. Cronbach alpha value was calculated for internal consistency reliability test. The relationship between the total score and the scores of the factors of the scale was evaluated by Spearman correlation analysis. CFA was performed using IBM AMOS 23 package program. The reason for using CFA is to test whether there is a fit between the variables that play a role in determining the theoretical factors and the original variables that make up the factors determined by EFA (Özdamar 2013). While evaluating the CFA fit indices, χ^2/df, standardised root mean squared residual (SRMR), root mean square error of approximation (RMSEA), Comparative Fit Index (CFI), Goodness of Fit Index (GFI), Adapted Goodness of Fit (AGFI), Normed Fit Index (NFI) and Non-Normed Fit Index (NNFI) values were used. $p < 0.05$ was considered statistically significant.

2.7 | Ethical Considerations

The study received ethical approval from the Ethics Committee of Kayseri University with approval number 57/2023 dated September 11, 2023, and institutional permission was obtained from the hospital where the study was conducted. All participants were informed about the purpose of the study before data collection, and they were assured that the data would be kept confidential and used only for scientific purposes. Additionally, participants were asked to sign a written, informed consent form. Throughout the study, care was taken to adhere to the principles of the Helsinki Declaration.

3 | Results

3.1 | Characteristics of Participants

The descriptive characteristics of the 485 participants included in the study are presented in Table 1. Since the study used a stratified sampling method based on participants' professional characteristics, it can be observed that the numbers of nurses and physicians participating in the study are comparable. Among the participants, 62.5% were female, 58.8% were married, 81.6% did not have any chronic diseases and 53.0% worked in clinics. In addition, the average age of the individuals included in the study was 32.31 ± 5.90, the median value of the total working time in the institution was 4 (1–23) and the median value of the total working time in the profession was 7 (1–23).

TABLE 1 | Descriptive characteristics of participants.

Variables	Statistics
Job, *n* (%)	
Nurse	255 (52.6)
Physician	230 (47.4)
Age, (year)[a]	32.31 ± 5.90 (22–49)
Gender, *n* (%)	
Female	303 (62.5)
Male	182 (37.5)
Marital status, *n* (%)	
Married	285 (58.8)
Single	200 (41.2)
Chronic disease (Hypertension, diabetes, thyroid diseases etc.), *n* (%)	
There is	89 (18.4)
No	396 (81.6)
Total working time in the institution, (year)[b]	4 (1–23)
Total working time in the profession, (year)[b]	7 (1–23)
Department, *n* (%)	
Emergency service	63 (13.0)
Intensive care	165 (34.0)
Clinic	257 (53.0)

Abbreviations: %, percentage value; *n*, the number of participants.
[a]Mean ± standard deviation (min–max).
[b]Median (min–max).

3.2 | Reliability Analysis

Statistical data on the reliability coefficients of the scale are presented in Table 2. In the literature, a corrected item-total correlation value of > 0.40 is recommended (Devon et al. 2007). Therefore, items with values < 0.40 were removed from the study step by step, and only items with values > 0.40 were left in the study. Item 2 and Item 19, with an item-total correlation value of < 0.40, were removed from the scale, and the scale statistics were recalculated. In the second stage, the item-total correlation values were re-evaluated, and since the correlation coefficient of Item 3 was low (< 0.40), Item 3 was also removed from the scale. Statistics for the remaining items are given in the last step.

3.3 | Exploratory and Confirmatory Factor Analysis

According to ANOVA with Tukey's test for non-additivity analysis, the scale has a summable structure ($F = 21.401$, $p < 0.001$). Kaiser–Meyer–Olkin sampling adequacy was above 0.5 (KMO = 0.926) and Bartlett's test of sphericity was significant ($\chi^2 = 5049.779$; $p < 0.001$). KMO assesses whether the correlation structure between variables is sufficient for factor analysis. Bartlett's test of sphericity tests whether the correlation matrix between variables is a unit matrix. In other words, it evaluates whether there is a significant relationship between the variables. It is seen that the scale has a factorisable structure (Özdamar 2013). Principal component analysis (PCA) was applied as the factor extraction method and the varimax rotation

TABLE 2 | Reliability coefficients of the scale.

		First step		Last step	
	Mean ± SD	Corrected item-total correlation	Cronbach's alpha if item deleted	Corrected item-total correlation	Cronbach's alpha if item deleted
Item-1	3.43 ± 1.23	0.618	0.916	0.598	0.919
Item-2	2.95 ± 1.39	0.366	0.921	—	—
Item-3	3.12 ± 1.30	0.404	0.920	—	—
Item-4	3.11 ± 1.29	0.467	0.919	0.441	0.922
Item-5	3.36 ± 1.32	0.616	0.916	0.617	0.918
Item-6	3.05 ± 1.42	0.437	0.920	0.431	0.923
Item-7	3.11 ± 1.29	0.485	0.918	0.479	0.921
Item-8	3.20 ± 1.32	0.484	0.918	0.472	0.922
Item-9	3.29 ± 1.29	0.689	0.914	0.688	0.917
Item-10	3.41 ± 1.24	0.750	0.913	0.757	0.915
Item-11	3.58 ± 1.18	0.715	0.914	0.720	0.916
Item-12	3.35 ± 1.29	0.648	0.915	0.655	0.917
Item-13	3.27 ± 1.29	0.641	0.915	0.646	0.918
Item-14	3.30 ± 1.36	0.658	0.915	0.670	0.917
Item-15	3.38 ± 1.32	0.667	0.915	0.685	0.917
Item-16	3.39 ± 1.25	0.731	0.914	0.748	0.915
Item-17	3.54 ± 1.30	0.630	0.916	0.640	0.918
Item-18	3.82 ± 1.27	0.594	0.916	0.607	0.919
Item-19	3.33 ± 1.31	0.335	0.921	—	—
Item-20	3.80 ± 1.29	0.564	0.917	0.573	0.919
Item-21	3.46 ± 1.33	0.447	0.919	0.436	0.923
Item-22	3.59 ± 1.26	0.502	0.918	0.513	0.921

method was applied, and rotated factor loadings were obtained in Table 3. PCA is a statistical technique used to reduce variables in multivariate data sets to a smaller number of new variables (components). When the line graph based on the eigenvalues of the factors in Table 3 and Graph 1 is analysed, it is seen that the 19 items consist of three components. The total variance explained by the three components is 58.927%. Rotated factor loadings obtained as a result of the varimax rotation method are given in Table 3. Cronbach's alpha values were 0.921 for the first factor, 0.766 for the second factor, 0.797 for the third factor and 0.923 for the total scale. As a result of CFA, χ^2/df, SRMR, RMSEA, CFI, GFI and NFI values, which are among the fit indices of the model, are given in Table 3. Table 3 shows the acceptable limits of fit. Considering these fit limits, the path diagram of the model obtained by creating three factors and the path diagram results with standardised coefficients are given in Figure 2. When the model fit indices were examined within the limits of fit in Table 3, acceptable fit was obtained in terms of RMSEA and χ^2/df, and values close to the threshold limit were obtained when examined in terms of other fit criteria, and acceptable fit was found when the scale was evaluated in terms of model fit indices (Harrington 2009; Kline 2023; Marsh et al. 2006; Schermelleh-Engel et al. 2003).

3.4 | Relationship Between the Total Scale Score and Subscales

The correlation coefficients between the total score of the scale and the factors are given in Table 4. A strong positive relationship was found between Factor 1 and Factor 2 and the total score. A moderate positive significant relationship was found between Factor 3 and the total score.

3.5 | Reliability Analysis

According to Table 5, there is no statistically significant difference between the first and second measurements for Factor-1, Factor-2, Factor-3 and total scores ($p > 0.05$). There is a high level of agreement between the first and second measurements in ICC values. In addition, the Cronbach alpha value of the scale was

TABLE 3 | Summary results of exploratory factor analysis and confirmatory factor analysis.

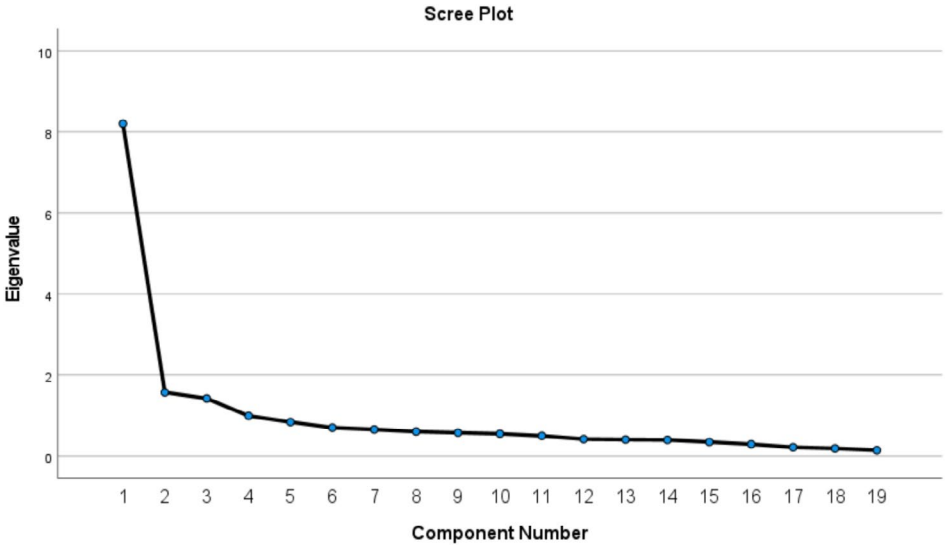
		Factor 1	Factor 2	Factor 3	Eigenvalues	% of variance explained	Cronbach alpha
Exploratory factor analysis results							
Item-15	I am concerned about making a medical error due to sleepiness during the shift.	**0.850**	0.125	0.142	8.206	43.190	0.921
Item-14	I am concerned about the negative impact of sleepiness on patient safety during the shift.	**0.846**	0.108	0.142			
Item-16	I am concerned that my sleepiness at the shift may affect my occupational health and safety.	**0.805**	0.172	0.276			
Item-13	I have difficulty in thinking/making decision process due to sleepiness at the shift.	**0.695**	0.316	0.078			
Item-12	The sleepiness I experience during the shift has a negative effect on my work performance.	**0.679**	0.370	0.065			
Item-10	There are moments when my concentration decreases during the shift due to the sleepiness I feel.	**0.632**	0.426	0.288			
Item-17	The sleepiness I experience during the shift has a negative effect on my daily activities during the shift.	**0.614**	0.184	0.340			
Item-11	There are times when I feel drowsy at the shift due to sleepiness.	**0.567**	0.366	0.381			
Item-9	I feel that my activities slow down during the shift because of the sleepiness I feel.	**0.540**	0.463	0.252			
Item-6	I have difficulty staying awake during patient handover at the end of my shift.	0.109	**0.743**	0.020	1.571	8.268	0.766
Item-7	My sleepiness intensity increases due to environmental factors (noise, light, heat, etc.) during the shift work.	0.144	**0.662**	0.163			
Item-4	I have difficulty staying awake during my free time at the shift.	0.188	**0.599**	0.086			
Item-8	I worry about losing my struggle to stay awake during the shift.	0.227	**0.598**	0.101			
Item-5	I try very hard not to fall asleep towards the end of my shift.	0.249	**0.577**	0.385			
Item-1	I often experience sleepiness during the shift.	0.382	**0.436**	0.316			
Item-22	I take preventive measures to avoid falling asleep during the shift, such as communicating with other staff or patients' relatives, changing the environment, washing my face or listening to music.	0.167	0.112	**0.815**	1.419	7.468	0.797
Item-21	I eat a snack to avoid falling asleep during the shift.	0.252	0.152	**0.771**			
Item-20	I take stimulants (tea, coffee, nicotine, etc.) to avoid falling asleep during the shift.	0.070	0.145	**0.760**			
Item-18	The sleepiness I experience during the shift has a negative effect on my social life after the shift.	0.449	0.146	**0.569**			
Total						58.927	0.923

(Continues)

TABLE 3 | (Continued)

Confirmatory factor analysis results						
Fit ındices	**X^2/df (p)**	**RMSEA**	**SRMR**	**CFI**	**GFI**	**NFI**
Reference value	$2 \leq \chi^2/df < 3$	0.05–0.08	0.05–0.10	0.90–0.95	0.90–0.95	0.90–0.95
Model	2.843	0.062	0.0534	0.948	0.922	0.922

Note: Factor 1: cognitive and physical effects of sleepiness on the process of working, Factor 2: difficulty keeping awake, Factor 3: sleepiness coping strategies. Extraction method: principal component analysis, Rotation method: Varimax with Kaiser normalisation, rotation converged in 5 iterations. The bold values in the table indicate the highest factor loadings for each item and demonstrate that the model has an acceptable level of fit.
Abbreviations: X^2/df, chi-squared/degree of freedom; CFI, comparative fit index; GFI, goodness of-fit; NFI, normed fit index; RMSEA, root means square error of approximation; SRMR, standardised root mean squared residual.



GRAPH 1 | Scree plot based on eigenvalues of factors.

found to be 0.852 in the test–retest analysis. According to these findings, it is seen that the scale is reliable.

4 | Discussion

It is crucial to use valid, reliable and standardised measurement tools to determine the levels of sleepiness among nurses and physicians working night shifts. However, it is observed that there is no specifically developed scale for nurses and physicians in the current literature. Considering this gap in the relevant literature, researchers therefore resolved that it would be appropriate to develop a scale specific to this field. It is believed that this research will contribute to the existing literature and shed light on descriptive and experimental studies regarding the sleepiness status of nurses and physicians working night shifts. As a result of the data analysis, a scale consisting of three factors and 19 items has been developed to measure the levels of sleepiness among nurses and physicians.

Content validity is defined as the degree to which each item and the entirety of the scale represent the intended characteristic successfully. This assessment is typically conducted in the context of expert opinions (Younas and Porr 2018). In the research, the Lawshe technique was used to assess content validity. In content validity calculations, it is expected to have a qualified expert number between 5 and 40 in order to obtain objective results (Ayre and Scally 2014; Morgan et al. 2019). In the study, expert opinions were obtained from 21 professionals in the field regarding the draft scale items. Additionally, within the framework of the Lawshe technique, it is emphasised that with the participation of 21 experts, the minimum CVR corresponding to an $\alpha = 0.05$ significance level should be 0.42 for scale items (Yesilyurt and Capraz 2018). In the scope of the study, due to the CVR being at least 0.71 and the Content Validity Index being calculated as 0.96 after expert opinions on the scale items, it was considered that there was a consensus among experts and content validity was ensured.

Item analysis is an analytical method used to understand the relationship between test items and the total test score. This analysis evaluates the contribution of each test item to the overall test performance, thereby revealing the internal structure of the measurement instrument (Morgan et al. 2019). In the literature, it is recommended that the corrected item-total correlation value be >0.40 (Beavers et al. 2019). Accordingly, three items with item-total score correlations below 0.40 were removed from the scale. It was observed that the remaining 19 items had item-total score correlations ranging from 0.43 to 0.75, which are statistically significant. The positive and significant correlation of the item-total score indicates that the items in the measurement tool sufficiently measure similar behaviours (Pallant 2020). When the overall item-total correlation is 0.40 or higher in particular, it is emphasised that the relevant items have the ability to effectively discriminate individuals (Beavers et al. 2019). This

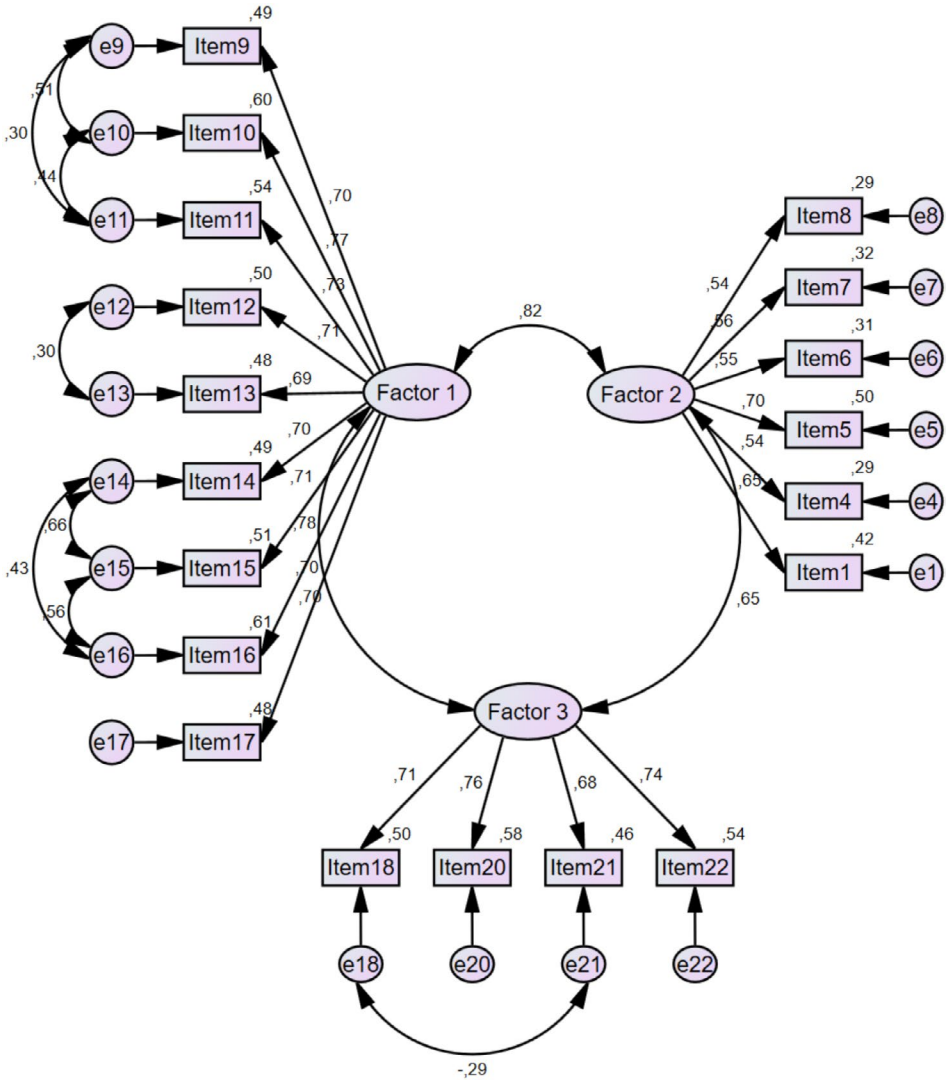


FIGURE 2 | Path diagram (Standardised estimates).

TABLE 4 | The relationship between the total scale score and factor scores.

Factors	Correlation coefficient, *p*	Total	Factor 1	Factor 2
Factor 1	Rho	0.915**	—	—
	p	<0.001	—	—
Factor 2	Rho	0.818**	0.613**	—
	p	<0.001	<0.001	—
Factor 3	Rho	0.663**	0.472**	0.442**
	p	<0.001	<0.001	<0.001

Note: Factor 1: cognitive and physical effects of sleepiness on the process of working, Factor 2: difficulty keeping awake, Factor 3: sleepiness coping strategies; rho, Spearman correlation coefficient.
**All values in the table are statistically significant at the 0.01 level.

indicates that the items of the scale measure reliably and validly. The item correlation coefficients in this study being above 0.40 demonstrate that the scale items possess discriminatory properties.

Factor analysis is a multivariate statistical method that aims to uncover the underlying structures behind an observed set of variables. It seeks specifically to reduce the variance among correlated variables to a smaller number of fundamental factors (Karaman et al. 2017). First, the Kaiser–Meyer–Olkin and Bartlett tests were performed to evaluate whether the sample was sufficient for factor analysis. For a good factor analysis, the Kaiser–Meyer–Olkin value is expected to be >0.80, and in this study, it was determined to be 0.926 (Williams et al. 2010). This value shows that the sample size is sufficient for factor analysis. Additionally, the chi-square value of the study's Bartlett test was statistically significant ($p<0.001$), indicating that this result provides a suitable basis for factor analysis.

Rotation of the axes in the factor analysis technique allows the clarity and significance of the items to be investigated, and it is expected that the explained variance ratio for multidimensional scales should be at least 30% (Pallant 2020). In this study, it was seen that 19 items, which explained 58.927% of the total variance with the rotation process, were grouped into three factors. The variance of the first factor was 43.190%, the variance of the second factor was 8.268% and the variance of the third factor was 7.468%. In scale development studies, it is emphasised that

TABLE 5 | Test–retest results ($n = 51$).

Factor	First measurement, Mean ± SD, Median (IQR)	Second measurement, Mean ± SD, Median (IQR)	*Z*	*p*	ICC	*p*
Factor 1	36.76 ± 7.12 38 (10)	37.12 ± 5.69 39 (6)	1.103	0.270	0.928 (0.877–0.958)	< 0.001
Factor 2	22.47 ± 5.30 23 (7)	22.76 ± 4.95 23 (7)	1.497	0.134	0.969 (0.972–0.991)	< 0.001
Factor 3	16.78 ± 3.25 18 (5)	17.06 ± 2.73 18 (4)	1.640	0.101	0.922 (0.868–0.955)	< 0.001
Total	76.02 ± 12.82 77 (18)	76.94 ± 10.59 79 (15)	1.792	0.073	0.958 (0.927–0.976)	< 0.001

Note: Factor 1: cognitive and physical effects of sleepiness on the process of working, Factor 2: difficulty keeping awake, Factor 3: sleepiness coping strategies.
Abbreviations: ICC, intraclass correlation coefficient; *Z*, Wilcoxon Signed Ranks test.

40% to 60% of the total variance of factor loadings is sufficient to explain, and increasing this ratio further increases the power (Carpenter 2018). As a result of the rotation process performed in this study, it was determined that 28 items, explaining 58.927% of the total variance, were grouped into three factors. When the items under the factors of the scale were examined, the first factor was named as 'cognitive and physical effects of sleepiness on the process of working', the second factor as 'difficulty keeping awake' and the third factor as 'sleepiness coping strategies'. The Cronbach's alpha values of the scale (0.923) and the factors (respectively: 0.921, 0.766, 0.797) were quite high. Considering that the reliability coefficient is above 0.70 in the literature, it can be said that the reliability coefficient of the total items and factors of the developed scale is quite high and the scale is reliable (Pallant 2020).

Confirmatory factor analysis is used to verify the fit of the model with the factors and is performed to determine whether the factors determined to evaluate the construct validity of the draft scale in scale development studies are statistically confirmed (Brown 2015; Kartal and Bardakçı 2018). As a result of the CFA, the fit indices of the model (χ^2/df, *p*, RMSEA, SRMR, CFI, GFI and NFI) were examined by considering the reference ranges accepted in the literature (Harrington 2009; Kline 2023; Marsh et al. 2006; Schermelleh-Engel et al. 2003). As a result of the analyses, it is seen that there is an acceptable fit in terms of RMSEA and χ^2/df criteria. When other fit criteria were analysed, values close to the threshold limit were obtained. When the scale was evaluated in terms of model fit indices, an acceptable fit was found.

One of the methods frequently used in reliability analysis is the test–retest method (Yaslioglu 2017). In this method, scale items are applied to the same individuals after a certain time interval and the correlation between the two measurements is expected not to fall below 0.70 (Morgan et al. 2019). In this study, the test–retest method was also applied for reliability analysis, and the calculations revealed that the correlation value for the first factor was 0.928, for the second factor it was 0.969, for the third factor it was 0.922 and for the total item correlation, it was 0.958. It was determined that all items and factor scores were significant between the two administrations ($p < 0.001$), and it was shown that the scale can be used reliably.

Scale items can be assessed using binary, ternary, quaternary or quintenary Likert-type scales. These ratings are selected based on the objectives of the data analysis processes (Younas and Porr 2018). In this study, a 5-point Likert scale was used for assessment. The rating scale was as follows: 1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree. There is no reverse-coded items in the scale. The minimum score for the first factor is 9 and the maximum score is 45, the minimum score for the second factor is 6 and the maximum score is 30, and the minimum score for the third factor is 4, whereas the maximum score is 20. The minimum score for all scale items is 19 and the maximum score is 95. An increase in the total score indicates higher severity of sleepiness in individuals.

4.1 | Limitations

Among the limitations of the study, it can be indicated firstly that focusing on sampling in a specific region may limit the ability to generalise. The fact that the data is limited to nurses and physicians working night shifts at a university hospital may limit the extension of the findings to a general population. In addition, relying on the subjective experiences of the participants in the development process of the scale may have a subjective impact on the general usability of the scale. The fact that the study was developed and implemented in a specific language and culture may create a limitation that it cannot cover different language and cultural groups. These limitations should be taken into account in interpreting the results of the study and assessing their general validity.

5 | Conclusion

In this study, a customised sleepiness scale was developed for nurses and physicians working night shifts. The Sleepiness Scale developed is suitable for determining the severity of sleepiness experienced by nurses and physicians during night shifts, and it can be used in related research. The findings of the study show that the scale, which consists of 19 items and three factors, is a valid and reliable measurement tool. It is believed that the scale can help in understanding the sleepiness status of nurses and physicians working night shifts in greater detail and in planning necessary interventions. The findings indicate that

the scale is an appropriate and effective tool for its purpose, and it has the potential to increase awareness of sleepiness among nurses and physicians working night shifts, as well as contribute to research in this area. Additionally, it is anticipated that the scale could provide a new perspective to research in the field of sleep. It is believed that evaluating its validity and reliability with nurses and physicians from different countries would contribute to the development of the scale.

Author Contributions

Ali Kaplan: conceptualisation, methodology, validation, formal analysis, investigation, resources, writing – original draft, writing – review and editing, visualisation, supervision, project administration. **Özlem Kaplan:** conceptualisation, methodology, validation, investigation, writing – original draft. **Cevriye Özdemir:** conceptualisation, methodology, validation, investigation, writing – original draft. **Ferhan Elmalı:** conceptualisation, methodology, formal analysis, writing – review and editing. **Emre Bülbül:** conceptualisation, methodology, validation, investigation, writing – original draft. The authors have checked to make sure that our submission conforms as applicable to the Journal's statistical guidelines described here.

Acknowledgements

The authors thank all the nurses and physicians who participated in this study.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ayre, C., and A. J. Scally. 2014. "Critical Values for Lawshe's Content Validity Ratio: Revisiting the Original Methods of Calculation." *Measurement and Evaluation in Counseling and Development* 47, no. 1: 79–86. https://doi.org/10.1177/0748175613513808.

Barnes, C. M., and N. F. Watson. 2019. "Why Healthy Sleep Is Good for Business." *Sleep Medicine Reviews* 47: 112–118. https://doi.org/10.1016/j.smrv.2019.07.005.

Beavers, A. S., J. W. Lounsbury, J. K. Richards, S. W. Huck, G. J. Skolits, and S. L. Esquivel. 2019. "Practical Considerations for Using Exploratory Factor Analysis in Educational Research." *Practical Assessment, Research & Evaluation* 18, no. 1: 6. https://doi.org/10.7275/qv2q-rk76.

Brown, T. A. 2015. *Confirmatory Factor Analysis for Applied Research*. Guilford Publications.

Carpenter, S. 2018. "Ten Steps in Scale Development and Reporting: A Guide for Researchers." *Communication Methods and Measures* 12, no. 1: 25–44. https://doi.org/10.1080/19312458.2017.1396583.

Devon, H. A., M. E. Block, P. Moyle-Wright, et al. 2007. "A Psychometric Toolbox for Testing Validity and Reliability." *Journal of Nursing Scholarship* 39, no. 2: 155–164. https://doi.org/10.1111/j.1547-5069.2007.00161.x.

Flo, E., B. Bjorvatn, S. Folkard, et al. 2012. "A Reliability and Validity Study of the Bergen Shift Work Sleep Questionnaire in Nurses Working Three-Shift Rotations." *Chronobiology International* 29, no. 7: 937–946. https://doi.org/10.3109/07420528.2012.699120.

Geiger Brown, J., M. Wieroney, L. Blair, et al. 2014. "Measuring Subjective Sleepiness at Work in Hospital Nurses: Validation of a Modified Delivery Format of the Karolinska Sleepiness Scale." *Sleep and Breathing* 18: 731–739. https://doi.org/10.1007/s11325-013-0935-z.

Gusman, E., J. Standlee, K. J. Reid, and L. F. Wolfe. 2023. "Work-Related Sleep Disorders: Causes and Impacts." *Seminars in Respiratory and Critical Care Medicine* 44: 385–395.

Härmä, M., K. Karhula, S. Puttonen, et al. 2019. "Shift Work With and Without Night Work as a Risk Factor for Fatigue and Changes in Sleep Length: A Cohort Study With Linkage to Records on Daily Working Hours." *Journal of Sleep Research* 28, no. 3: e12658. https://doi.org/10.1111/jsr.12658.

Harrington, D. 2009. *Confirmatory Factor Analysis*. Oxford University Press.

Johns, M. W. 1992. "Reliability and Factor Analysis of the Epworth Sleepiness Scale." *Sleep* 15, no. 4: 376–381. https://doi.org/10.1093/sleep/15.4.376.

Karaman, H., A. Burcu, and D. Ç. Aktan. 2017. "The Comparison of Factor Extraction Methods Used in Exploratory Factor Analysis." *Gazi University Gazi Faculty of Education Journal* 37, no. 3: 1173–1193. https://doi.org/10.17152/gefad.309356.

Kartal, M., and S. Bardakçı. 2018. *SPSS ve AMOS Uygulamalı Örneklerle Güvenirlik ve Geçerlik Analizleri*. Akademisyen Kitabevi. https://doi.org/10.37609/akya.1623.

Kecklund, G., and J. Axelsson. 2016. "Health Consequences of Shift Work and Insufficient Sleep." *British Medical Journal* 355: 1–13. https://doi.org/10.1136/bmj.i5210.

Kline, R. B. 2023. *Principles and Practice of Structural Equation Modeling*. Guilford Publications.

Kyngäs, H. 2020. "Qualitative Research and Content Analysis." In *The Application of Content Analysis in Nursing Science Research*, 3–11. Springer International Publishing. https://doi.org/10.1007/978-3-030-30199-6_1.

Lee, M., E. Whitsel, C. Avery, et al. 2022. "Variation in Population Attributable Fraction of Dementia Associated With Potentially Modifiable Risk Factors by Race and Ethnicity in the US." *JAMA Network Open* 5, no. 7: e2219672. https://doi.org/10.1001/jamanetworkopen.2022.19672.

Leger, D., and C. Stepnowsky. 2020. "The Economic and Societal Burden of Excessive Daytime Sleepiness in Patients With Obstructive Sleep Apnea." *Sleep Medicine Reviews* 51: 101275. https://doi.org/10.1016/j.smrv.2020.101275.

Low, D. V., M. N. Wu, and A. P. Spira. 2019. "Sleep Duration and Cognition in a Nationally Representative Sample of US Older Adults." *American Journal of Geriatric Psychiatry* 27, no. 12: 1386–1396. https://doi.org/10.1016/j.jagp.2019.07.001.

Ma, Y., L. Liang, F. Zheng, L. Shi, B. Zhong, and W. Xie. 2020. "Association Between Sleep Duration and Cognitive Decline." *JAMA Network Open* 3, no. 9: e2013573. https://doi.org/10.1001/jamanetworkopen.2020.13573.

Maclean, A. W., G. C. Fekken, P. Saskin, and J. B. Knowles. 1992. "Psychometric Evaluation of the Stanford Sleepiness Scale." *Journal of Sleep Research* 1, no. 1: 35–39. https://doi.org/10.1111/j.1365-2869.1992.tb00006.x.

Marsh, H. W., K.-T. Hau, C. Artelt, J. Baumert, and J. L. Peschar. 2006. "OECD'S Brief Self-Report Measure of Educational Psychology's Most Useful Affective Constructs: Cross-Cultural, Psychometric Comparisons Across 25 Countries." *International Journal of Testing* 6, no. 4: 311–360. https://doi.org/10.1207/s15327574ijt0604_1.

Martin, V. P., R. Lopez, Y. Dauvilliers, J.-L. Rouas, P. Philip, and J.-A. Micoulaud-Franchi. 2023. "Sleepiness in Adults: An Umbrella Review

of a Complex Construct." *Sleep Medicine Reviews* 67: 101718. https://doi.org/10.1016/j.smrv.2022.101718.

Matricciani, L., Y. S. Bin, T. Lallukka, et al. 2017. "Past, Present, and Future: Trends in Sleep Duration and Implications for Public Health." *Sleep Health* 3, no. 5: 317–323. https://doi.org/10.1016/j.sleh.2017.07.006.

Morgan, G. A., K. C. Barrett, N. L. Leech, and G. W. Gloeckner. 2019. *IBM SPSS for Introductory Statistics: Use and Interpretation*. 6th ed. Routledge.

Nena, E., M. Katsaouni, P. Steiropoulos, E. Theodorou, T. C. Constantinidis, and G. Tripsianis. 2018. "Effect of Shift Work on Sleep, Health, and Quality of Life of Health-Care Workers." *Indian Journal of Occupational and Environmental Medicine* 22, no. 1: 29–34. https://doi.org/10.4103/ijoem.IJOEM_4_18.

Özdamar, K. 2013. *Statistical Data Analysis With Package Programs*. Nisan Bookstore.

Pallant, J. 2020. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*. McGraw-Hill education. https://doi.org/10.4324/9781003117452.

Reid, K. J., and S. M. Abbott. 2015. "Jet Lag and Shift Work Disorder." *Sleep Medicine Clinics* 10, no. 4: 523–535. https://doi.org/10.1016/j.jsmc.2015.08.006.

Romero Jeldres, M., E. Díaz Costa, and T. Faouzi Nadim. 2023. "A Review of Lawshe's Method for Calculating Content Validity in the Social Sciences." *Frontiers in Education* 8: 1271335. https://doi.org/10.3389/feduc.2023.1271335.

Schermelleh-Engel, K., H. Moosbrugger, and H. Müller. 2003. "Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-Of-Fit Measures." *Methods of Psychological Research Online* 8, no. 2: 23–74.

Smart, D., L. James, T. Odom-Maryon, and S. Rowan. 2020. "Effect of Subjective Sleepiness on Critical Skills Performance in National Guard Medical Personnel During Extensive Disaster Training." *Health Emergency and Disaster Nursing* 7, no. 1: 17–26. https://doi.org/10.24298/hedn.2019-0007.

South, L., D. Saffo, O. Vitek, C. Dunne, and M. A. Borkin. 2022. "Effective Use of Likert Scales in Visualization Evaluations: A Systematic Review." *Computer Graphics Forum* 41: 43–55.

Spilsbury, J. C., D. Drotar, C. L. Rosen, and S. Redline. 2007. "The Cleveland Adolescent Sleepiness Questionnaire: A New Measure to Assess Excessive Daytime Sleepiness in Adolescents." *Journal of Clinical Sleep Medicine* 3, no. 6: 603–612. https://doi.org/10.5664/jcsm.26971.

Streiner, D., G. R. Norman, and J. Cairney. 2016. "Health Measurement Scales: A Practical Guide to Their Development and Use." *Australian and New Zealand Journal of Public Health* 17, no. 3: 189. https://doi.org/10.1891/1521-0987.17.3.150.

Tabachnick, B. G., and L. S. Fidell. 2014. *Using Multivariate Statistics*. Allyn and Bacon.

Williams, B., A. Onsman, and T. Brown. 2010. "Exploratory Factor Analysis: A Five-Step Guide for Novices." *Australasian Journal of Paramedicine* 8: 1–13. https://doi.org/10.33151/ajp.8.3.93.

Xu, W., C.-C. Tan, J.-J. Zou, X.-P. Cao, and L. Tan. 2020. "Sleep Problems and Risk of All-Cause Cognitive Decline or Dementia: An Updated Systematic Review and Meta-Analysis." *Journal of Neurology, Neurosurgery & Psychiatry* 91, no. 3: 236–244. https://doi.org/10.1136/jnnp-2019-321896.

Yaslioglu, M. M. 2017. "Factor Analysis and Validity in Social Sciences: Application of Exploratory and Confirmatory Factor Analyses." *Istanbul University Journal of the School of Business* 46: 74–85.

Yesilyurt, S., and C. Capraz. 2018. "A Road Map for the Content Validity Used in Scale Development Studies." *Erzincan University Journal of Education Faculty* 20, no. 1: 251–264. https://doi.org/10.17556/erziefd.297741.

Younas, A., and C. Porr. 2018. "A Step-By-Step Approach to Developing Scales for Survey Research." *Nurse Researcher* 26, no. 3: 14–19. https://doi.org/10.7748/nr.2018.e1585.