Original Article / *Özgün Makale*

DOI: 10.4274/jtsm.galenos.2024.19327

*J Turk Sleep Med*

# *Development and Investigation into the Psychometric Properties of the Troy Sleep Scale*

## *Troya Uyku Ölçeği'nin Geliştirilmesi ve Psikometrik Özelliklerinin İncelenmesi*

Erhan Akıncı, Nurullah Bolat*, Şükrü Alperen Korkmaz**, Çetin Toraman***

*Acıbadem Ataşehir Hospital, Clinic of Psychiatry, İstanbul, Turkey*
**Çanakkale Onsekiz Mart University Faculty of Medicine, Department of Child and Adolescent Psychiatry, Çanakkale, Turkey*
***Çanakkale Onsekiz Mart University Faculty of Medicine, Department Psychiatry, Çanakkale, Turkey*
****Çanakkale Onsekiz Mart University Faculty of Medicine, Department of Medical Education, Çanakkale, Turkey*

### Abstract

**Objective:** A unique and comprehensive self-report scale to help assess sleep in Turkish is needed. This study aimed to develop the Troy Sleep Scale (TSS) and perform validity and reliability analyses.

**Materials and Methods:** The study was conducted with 494 medical students and residents. The psychometric properties of the devised scale were evaluated through exploratory and confirmatory factor analyses, Cronbach's alpha and McDonald's omega reliability evaluations, and item analyses based on both classical test theory and item response theory (IRT). The reliability assessment of the TSS was carried out through test-retest, and criterion-referenced validity was employed to provide additional validation.

**Results:** The TSS consists of 11 items with a 3-factor structure with eigenvalues greater than 2. The factors provided information about (I) dysfunction due to sleep problems, (II) affective symptoms due to sleep problems, and (III) sleep quality. The reliability levels of all three sub-factors and overall scale scores were 0.82 and above. The test-retest reliability level had a correlation value of >0.80. The item analyses conducted in line with the IRT showed that the scale items provided a high level of information and functioned together with the answer set. The correlation between the TSS and the Pittsburgh Sleep Quality Index was analyzed for criterion-referenced validity, yielding a correlation of 0.73.

**Conclusion:** The psychometric properties of the TSS indicated that it was valid and reliable, making it suitable for use as a screening test for individuals with sleep-related complaints.

**Keywords:** Sleep, sleep disorder, scale development, Turkish scale, validity, reliability

### Öz

**Amaç:** Türkçe'de uykuyu değerlendirmeye yardımcı olacak özgün ve kapsamlı bir öz bildirim ölçeğine ihtiyaç vardır. Bu çalışmanın amacı, Troya Uyku Ölçeğini (TUÖ) geliştirmek, geçerlilik ve güvenilirlik analizlerini yapmaktır.

**Gereç ve Yöntem:** Çalışma 494 tıp öğrencisi ve asistan hekim ile yürütülmüştür. Geliştirilen ölçeğin psikometrik özellikleri, açımlayıcı ve doğrulayıcı faktör analizleri, Cronbach alfa ve McDonald's omega güvenilirlik değerlendirmeleri ve hem klasik test teorisi hem de madde tepki teorisine (MTK) dayalı madde analizleri ile değerlendirilmiştir. TUÖ'nin güvenilirlik değerlendirmesi test-tekrar test yoluyla gerçekleştirilmiş ve ek doğrulama sağlamak için ölçüt referanslı geçerlilik kullanılmıştır.

**Bulgular:** TUÖ, özdeğeri 2'den büyük üç faktörlü bir yapıya sahip 11 maddeden oluşmaktadır. Faktörler, (I) uyku sorunlarına bağlı işlev bozukluğu, (II) uyku sorunlarına bağlı duygudurumu belirtileri ve (III) uyku kalitesi hakkında bilgi sunmaktadır. Her üç alt faktörün ve genel ölçek puanlarının güvenilirlik düzeyleri 0,82 ve üzerindedir. Test-tekrar test güvenirlik düzeyi >0,80 korelasyon değerine sahiptir. MTK doğrultusunda yapılan madde analizleri, ölçek maddelerinin yüksek düzeyde bilgi sağladığını ve cevap seti ile birlikte işlev gördüğünü kanıtlamıştır. TUÖ ile Pittsburgh Uyku Kalitesi İndeksi arasındaki korelasyon, ölçüt referanslı geçerlilik için analiz edilmiş ve 0,73'lük bir korelasyon elde edilmiştir.

**Sonuç:** TUÖ'nün psikometrik özellikleri, geçerli ve güvenilir olduğunu ve uyku ile ilgili yakınmaları olan bireyler için bir tarama testi olarak kullanılmaya uygun olduğunu göstermiştir.

**Anahtar Kelimeler:** Uyku, uyku bozukluğu, ölçek geliştirme, Türkçe ölçek, geçerlik, güvenirlik

**Address for Correspondence/Yazışma Adresi:** Erhan Akıncı, M.D., Assoc. Prof., Private Physician, Psychiatry Specialist, İstanbul, Turkey
**E-mail:** drerhanakinci@yahoo.com **ORCID-ID:** orcid.org/0000-0003-3331-8165
**Received/Geliş Tarihi:** 31.07.2024 **Accepted/Kabul Tarihi:** 12.09.2024 **Epub:** 12.02.2025

**Cite this article as:** Akıncı E, Bolat N, Korkmaz ŞA, Toraman Ç. Development and investigation into the psychometric properties of the troy sleep scale. Journal of Turkish Sleep Medicine.

## Introduction

Sleep is a fundamental requirement for human life. Regular and consistent sleep is crucial for maintaining overall health and well-being. Sleep is a dynamic process involving intricate neurochemical activity characterized by periodic transitions between different sleep stages. The structure and duration of sleep are influenced by various factors such as age and genetic predisposition.[1] Insomnia, a common symptom, often disappears when the underlying cause is addressed. Chronic insomnia is believed to affect approximately 10% of the population.[2,3] Insomnia is generally defined as the inability to initiate and maintain sleep which impairs sleep quality and integrity despite favorable conditions.[4] The negative impact of poor sleep quality is evident during the day, causing disability, impairment of cognitive and emotional functions, and behavioral problems. Sleep deprivation and fatigue often lead to daytime sleepiness, mood disturbances, attention and memory impairment, and an increased likelihood of accidents and errors.[5] Sleep problems are a common health problem in university students, insomnia is seen at a higher rate than in the general population in systemic reviews (18.5% vs. 7.5%), and sleep problems cause medical problems and stress in students and negatively affect their academic performance.[6-9] According to subjective experiences, sleep quality can be defined as satisfaction with the previous night's sleep and satisfactory wakefulness the next day. However, when subjective sleep experiences and evaluations cannot be measured or expressed numerically, they only have descriptive features. Scientific research can be conducted using numerical data obtained with valid and reliable measurement tools, and measurements made with sensitive measurement tools make the research valuable. For this purpose, scales facilitate the measurement of variable characteristics and determine the quality of the results.[10] Objective measurements such as polysomnography and actigraphy are costly and challenging for patients with sleep-related complaints.[11] Self-report methods, such as sleep diaries, sleep recording, and sleep scales, are commonly used to assess sleep quality, although they do not offer information about sleep structure. These methods attempt to gauge both the quantitative and qualitative aspects of sleep. The components and importance of sleep quality differ among individuals. Therefore, a self-report-based assessment is necessary to measure sleep quality.[12] In addition, self-report scales have many advantages, such as being inexpensive, practical, and quick to administer. Many scales have been developed to assess sleep problems and are used in Turkey (see review).[13] Most scales for sleep problems originate from abroad, and Turkish validity and reliability studies have been conducted.[14-19] In Turkey, there is a need for a local and comprehensive self-report scale to help distinguish individuals with sleep problems from healthy individuals and to determine sleep problems qualitatively and quantitatively. Therefore, this study aimed to develop a comprehensive and reliable self-report scale, the Troy Sleep Scale (TSS), to assess adult sleep quality.

## Materials and Methods

### Participants

The study was conducted between January 2022 and August 2022 with students studying at the Çanakkale Onsekiz Mart University Faculty of Medicine during the 2021-2022 academic year. The data were obtained from 494 medical students. The mean age of the students was 21.4±3.0 years. Of the participating students, 181 (36.6%) were male and 313 (63.4%) were female. Eighty-seven (17.6%) were class 1.144 (29.1%) were class 2.63 (12.8%) were class 3.40 (8.1%) were class 4.94 (19%) were class 5.35 (7.1%) were class 6 (intern doctors), and 31 (6.3%) were residents. Data from 494 medical students were randomly divided into two groups. Hair et al.[20] stated that exploratory factor analysis (EFA) results should be validated in a split sample from the original dataset or a separate sample obtained with a new application. The larger group comprised approximately 59% of the data (n=289) and was utilized for EFA, Cronbach's alpha, and McDonald's reliability coefficient calculations. The remaining 197 medical students' data were employed for confirmatory factor analysis (CFA). All 494 students' data were utilized for item response theory (IRT).

### Process

### Drafting of the Troy Sleep Scale

The TSS draft form was prepared as follows: (I) The scale aimed to evaluate adult sleep. (II) It was decided to use 5-degree Likert-type items (never, rarely, sometimes, most of the time, and always) as the answer set for the scale. Studies show that the options for the 5-point Likert structure work well.[21] (III) A literature review was conducted on scales assessing sleep, and candidate items were created by reviewing the relevant scales. (IV) The candidate item pool was presented to four expert psychiatrists working on sleep and one expert academician for measurement and evaluation. The experts kept items deemed appropriate on the trial form and removed those deemed inappropriate. While there were 29 items in the item pool, 23 remained after expert evaluation. (V) For Turkish language comprehensibility and plain expression, the opinion of an expert in Turkish was considered. The Turkish language expert evaluated the remaining 23 items and suggested ensuring more straightforward expressions. The items were revised in line with the recommendations of the linguist. Following the abovementioned process, a draft trial form consisting of 23 Likert-type items with a 5-point response set was obtained. Ten students participated in a pilot study using the draft form. This pilot study evaluated whether the scale items and expressions were understandable for the target group. The pilot study showed that the items and expressions were understandable.

Several scales have been developed to evaluate sleep quality, with the Pittsburgh Sleep Quality Index (PSQI) being the most commonly used.[22] The PSQI, which comprises 19 questions and seven dimensions, assesses subjective sleep quality, sleep latency, sleep duration, sleep efficiency, sleep disturbance, sleep medication use, and daytime dysfunction. A validity and reliability study of the PSQI was conducted by Ağargün et al.[14]

in Turkey, highlighting potential limitations of the scale, such as subjective interpretations affecting scores for dimensions like sleep duration, latency, and efficiency. In this study, the PSQI was administered to all participants.

### After the Development of the Draft Scale

The research was completed as follows. The approval of Çanakkale Onsekiz Mart University Clinical Research Ethics Committee (approval number: 2021-07, date: 20.10.2021) was obtained. Before the study, a consent form was obtained from all participants, indicating their agreement to participate. A draft form of the TSS was also prepared, and a trial application was carried out to obtain validity and reliability evidence of the scale. The trial applications were conducted face-to-face. To assess the validity of the TSS, item-total correlations were examined based on the classical test theory (CTT), which constructs validity within the scope of factor analysis and CFA. The scale's validity was further assessed through item discrimination and difficulty levels, item characteristic curves, and item and test information functions based on the IRT. The scale's reliability was examined using Cronbach's alpha internal consistency coefficient, McDonald's composite reliability coefficient, and marginal reliability coefficients.

### Statistical Analysis

### Data File Preparation Phase

The study involved analyzing forms completed by participants. It was observed that all data were present on the participant's forms. The data were subsequently transferred to JAMOVI and R statistical software. The "mvn" package of R was utilized to examine the multivariate normal distribution of the 23 items using the "henze-zirkler," "mardia," and "doornik-hansen" methods.[23] The Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity were employed to analyze the data. The KMO value was interpreted based on the reference values provided. Specifically, values above 0.89 (0.90 to 1.00) were considered "excellent," values between 0.80 and 0.89 were deemed "good," values between 0.70 and 0.79 were labeled "acceptable," values between 0.60 and 0.69 were categorized as "moderate," values between 0.50 and 0.59 were classified as "low level," and values below 0.50 were deemed "unacceptable".[24] In Bartlett's test of sphericity, the hypothesis $H_0$ was evaluated to determine whether the correlation matrix was a unit matrix, indicating the absence of relationships between the items.

### Factor Identification with Exploratory Factor Analysis (EFA)

Factor analyses were performed utilizing the principal axis factoring (PAF) methodology. PAF analyzes the shared variance among the measures instead of the error sources unique to each one. PAF, commonly used in social and behavioral science research, models the shared variance in a series of X-measures.[25] In determining the number of EFA factors, the reference value of "eigenvalue" was accepted as "1". A meaningful discussion in EFA involves assessing the extent to which the total variance (variability) of the characteristic being measured through the scale is explained. The structure obtained from factor analysis is based on Hair et al.[20] who states that in social sciences, where information is generally less precise, a solution that explains 60% of the total variance is acceptable. Warner[25] suggests that the acceptable limits are between 40% and 70%. In interpreting the factor structure, the minimum values for factor loading should be between ± 0.30 and ± 0.40. A factor-loading value of ± 0.50 is important, while ± 0.70 and above is indicative of a well-defined structure.[20]

### Determination of Reliability

Cronbach's alpha, McDonald's omega, and marginal reliability coefficients were used to determine the scale's reliability. A test-retest application was performed to determine the reliability level. Sixty-six participants were administered the scale again one month later. The level of consistency was examined using a correlation analysis between the scores of the first and last applications. The test-retest administration assessed the consistency of the measurement tool, whether it gave stable results over time, and provided evidence of reliability. Adequate reliability should be 0.70 and above, according to Nunnally and Bernstein[26].

### Confirmatory Factor Analysis (CFA)

In CFA analysis, the degree of model fit is important. By looking at the fit values, the suitability of the model can be determined. The literature suggests that reference values for fit indexes determined for CFA are acceptable for 0.05< root mean square error of approximation (RMSEA) ≤0.08 and excellent for 0>RMSEA≤0.05; a good fit for a Tucker-Lewis index (TLI) is 0.95 and above; and an acceptable fit for a comparative fit index (CFI) is 2< $X^2$/standard deviation (sd)≤5 and a good fit is 0>$X^2$/sd≤2.[27-34]

### Data Analyses with Item Response Theory (IRT)

IRT-based measurement tools offer several advantages, including the independence of item parameters and group characteristics. Additionally, IRT allows for unique standard error estimates for each participant. In IRT analyses, item parameters are independent of the participant group, and group characteristics are independent of the item sample.[16] Furthermore, standard errors can be estimated separately by analyzing the test results for each respondent. In this context, a standard framework for evaluating their ability according to IRT can be revealed, even if the respondents are tested with different questions.[35] The validity and reliability analyses using the IRT should include an examination of the unidimensionality and local independence assumptions.[36] Unidimensionality requires that only one characteristic is assessed (the relevant items of the measurement tool are only for one characteristic), which affects the performance of individuals on the measurement tool.[37] An item correlation matrix or EFA can be used to assess unidimensionality. This study used EFA to analyze unidimensionality, as explained in detail in the results section. According to the EFA results, the TSS had three subdimensions. In this case, each factor was accepted as unidimensional separately, and IRT analyses were performed accordingly. The assumption of local independence was tested using the Q3

statistic[38], and IRT calibrations were performed using the "Mirt v.1.30" package within the R v.4.1.2 program.[39] In the IRT, the discrimination value of an ideal scale item (i.e., "a" parameter) should be between 0.5 and 2. The parameter is generally within the acceptable range of 0.75 to 2.50 in the literature.[40] The ideal (medium difficulty level) limits for item difficulty levels (i.e., "b" parameter) are accepted between -1.00 and 1.00.[41] In aptitude or achievement tests, items with difficulty levels lower than -1.00 are considered easy, while items with difficulty levels higher than 1.00 are considered difficult.[42] The item information function is a graphical representation that illustrates the range of the trait (the trait being measured on the scale) in which the item best distinguishes individuals taking the measurement tool.[43]

## Results

### CTT Validity Evidence of the Troy Sleep Scale

#### Construct Validity (EFA)

The normal distribution of the 23 items on the scale was analyzed. Multivariate normal distribution examination revealed a multivariate normal distribution ($p>0.05$). The KMO value was calculated as 0.89. Bartlett's sphericity test value was calculated as 2040, and the result was significant ($df=55$, $p<0.05$). As explained in the data analysis section, these results demonstrate that the data are suitable for factor analysis. PAF was adopted for the EFA. According to the data of the 23 items in the EFA data file, item-total correlation values and factor analysis input loadings were analyzed. Twelve items (3, 4, 6, 7, 8, 10, 11, 12, 17, 18, 19, and 20) with initial, extraction, and item-total correlation values below 0.30 were identified. These items were removed from the scale. The results are presented in Table 1. The initial, extraction, and corrected item-total correlation values of the remaining 11 items on the TSS were between 0.45 and 0.79. As explained in the data analysis section, these values are within the ranges recommended in the literature for EFA. Eigenvalue and scree plot analyses were performed to determine the number of factors the TSS showed with the remaining 11 items. According to the eigenvalue data, these three factors had values greater than 1. The eigenvalue of the first factor was 2.66, and the explained variance was 24.2%. The eigenvalue of the second factor was 2.38, and the variance explained by it was 21.7%. The eigenvalue of the third factor was 2.24, and the explained variance was 20.4%. Together, the three factors explain 66.2% of sleep features. As described in the Data Analysis section, this value has been accepted in the literature. The scree plot obtained from the TSS confirmed a three-factor structure (Figure 1). The scree plot shows that the TSS had three factors, all of which had eigenvalues greater than 1. Based on these results, the scale was determined to have a three-factor structure. Axis rotation was performed to determine the factors in which the 11 items of the TSS were located. As three factors were determined to be related, oblimin rotation was performed. The items for these factors are listed in Table 2. Four items were grouped under factor 1, three under factor 2, and four under factor 3. After the rotation process, the factor load values ranged from 0.44 to 0.97. The items under these factors were analyzed, and the factors were named. Factor 1 consisted of the 9th, 14th, 15th, and 16th items and was named Dysfunction Due to Sleep Disorder (DSD) (minimum 4, maximum 20 points). Factor 2 consisted of items 21, 22, and

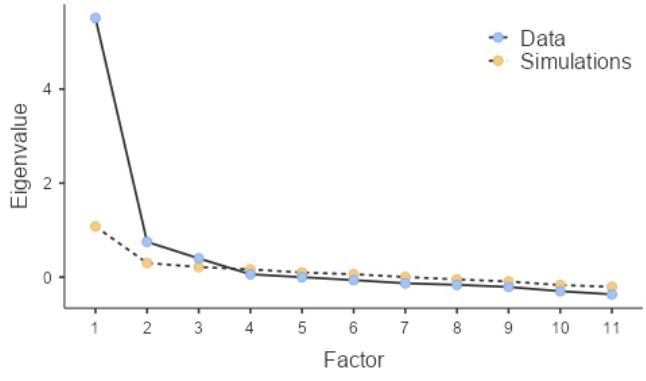


**Figure 1.** Scree plot obtained from Troy Sleep Scale data in exploratory factor analysis

**Table 1. EFA initial loadings and item-total correlations**

| Item | Initial | Extraction | Corrected item-total correlation |
|---|---|---|---|
| 1 | 0.68 | 0.77 | 0.71 |
| 2 | 0.58 | 0.58 | 0.58 |
| 5 | 0.49 | 0.50 | 0.46 |
| 9 | 0.46 | 0.45 | 0.63 |
| 13 | 0.50 | 0.52 | 0.65 |
| 14 | 0.63 | 0.64 | 0.76 |
| 15 | 0.72 | 0.61 | 0.74 |
| 16 | 0.76 | 0.70 | 0.79 |
| 21 | 0.63 | 0.66 | 0.66 |
| 22 | 0.68 | 0.68 | 0.68 |
| 23 | 0.62 | 0.65 | 0.69 |

EFA: Exploratory factor analysis

23 and was labeled Affective Symptoms Due to Sleep Disorder (ASD) (minimum 3, maximum 15 points). Factor 3 consisted of items 1, 2, 5, and 13 and was called sleep quality (SQ) (minimum 4, maximum 20 points). No reverse-scored items were used for these factors.

### Construct Validity (CFA)

The analyses revealed a 3-factor structure for the TSS, which consisted of 11 Likert-type items. The accuracy of this structure was analyzed using CFA. The diagram obtained from the analysis of the CFA data file is shown in Figure 2. When the diagram obtained from the CFA was examined, the 5th item under the 3rd factor had the lowest correlation (0.52). The highest correlation value was the 22nd item under the 2nd factor. As a result of the CFA, the fit indices were calculated as $X^2$/sd=1.98, CFI=0.98, TLI=0.97, SRMR=0.04, and RMSEA=0.06. In the DFA analysis, a covariance link was established by modifying the error terms of items 1-2 and 15-16. In this case, a correlation between the error terms of items 1 and 2 as well as 15 and 16 was accepted. Creating a covariance modification between the unknown parts of these items also showed that there were some semantic links between the items.

### Reliability of Troy Sleep Scale

The reliability of the TSS was analyzed using Cronbach's alpha, McDonald's omega, and the marginal reliability coefficients. The 11 items were subjected to reliability analyses as a single-factor scale with three separate factors. Cronbach's alpha reliability value of the first factor was 0.89, McDonald's omega reliability value was 0.89, and the marginal reliability value was 0.88. The Cronbach's alpha reliability value of the second factor was 0.88, the McDonald's omega reliability value was 0.89, and the marginal reliability value was 0.86. The Cronbach's alpha reliability value of the third factor was 0.82, McDonald's omega reliability value was 0.83, and the marginal reliability value was 0.89. The Cronbach's alpha reliability value of the TSS was 0.91 and the McDonald's omega reliability value was 0.92. The values obtained from the analyses of the reliability levels of the scale are 0.70 and above and had sufficient reliability.[15] A test-retest procedure was performed to determine the reliability level. Sixty-six participants were administered the scale again one month later. The level of consistency was examined by comparing the scores of the test and retest administrations. Table 3 presents the correlation results. The first-factor correlation of

**Table 2. Factors and items in the factors after oblimin rotation**

| Items | | Factors | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| **15** | The sleep problem I experience at night prevents me from doing my work during the day. | 0.972 | | |
| **16** | My night-time sleep problems impair my daytime functioning. | 0.811 | | |
| **9** | I cannot focus on my work during the day due to insomnia at night. | 0.525 | | |
| **14** | I suffer from forgetfulness during the day due to sleep problems at night. | 0.519 | | |
| **22** | I am constantly nervous and irritable due to sleep problems. | | 0.967 | |
| **21** | My tolerance for people is decreasing due to sleep problems. | | 0.771 | |
| **23** | I'm easily depressed because of the sleep problems I've been having. | | 0.714 | |
| **1** | I have trouble falling asleep at night. | | | 0.890 |
| **2** | I spend much time in bed at night until I fall asleep. | | | 0.870 |
| **13** | I have anxiety about not being able to sleep when I go to bed. | | | 0.479 |
| **5** | If I wake up at night, I find it difficult to fall asleep again. | | | 0.440 |

Factor 1: DSD: Dysfunction due to sleep disturbance, Factor 2: ASD: Affective symptoms due to sleep disturbance, Factor 3: SQ: Sleep quality

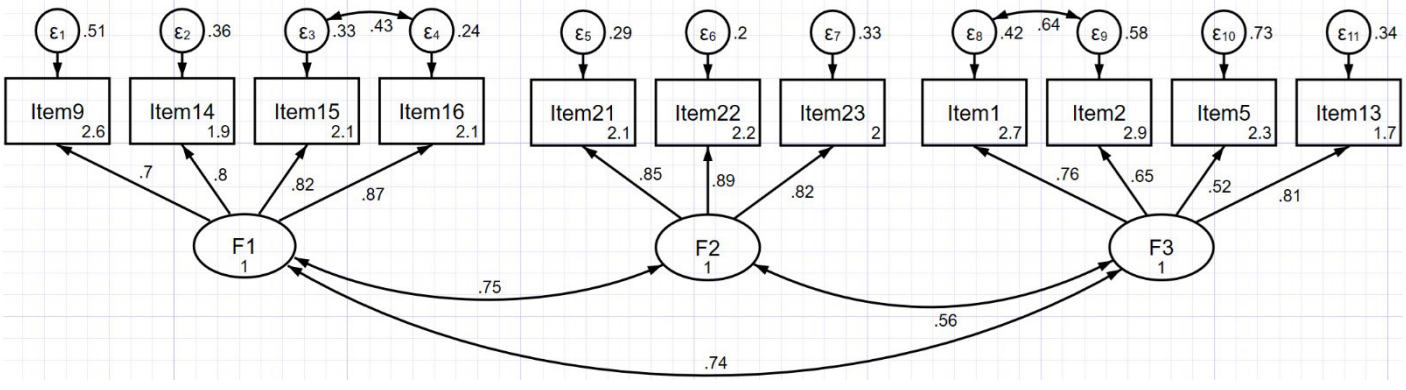


**Figure 2.** Confirmatory factor analysis diagram of the 3-factor structure of the Troy Sleep Scale

the scale, which was administered to 66 participants with an interval of one month and whose structure was discovered according to EFA, was 0.81, the second-factor correlation was 0.80, and the third-factor correlation was 0.86. For the 11-item total score, the correlation between the data from the first and second administration was analyzed, and a correlation value of 0.88 was obtained. The correlation values obtained correspond to high reliability.

### Validity Evidence for the Troy Sleep Scale

#### Factor 1 IRT Validity Evidence

IRT analyses were conducted using data from 494 students. The generalized partial credit model (GPCM) calibrated the IRT factor item. The GPCM calculated S_$\chi^2$ (degrees of freedom), RMSEA, level of significance, "a" (item discrimination) and "b" (item difficulty) parameters, and standard errors for each item separately (Table 4). According to the GPCM, the RMSEA threshold value for acceptable item fit is 0.08. With a value below this threshold, an item is considered to fit the model. The item fit statistics revealed that all items had RMSEA values below 0.05, indicating that the provided model fit according to the GPCM. The discrimination value, or "a" parameter, of an ideal scale item in IRT should be between 0.5 and 2. In the literature, the range for this parameter is between 0.75 and 2.50, which is considered acceptable. Based on the GPCM estimations (log-likelihood, $p<0.05$), the compatibility of the items was confirmed. The item characteristic curves depicted that Items 9, 14, 15, and 16 worked well with their options in the response set. The item information function is a graphical representation that illustrates the range of the feature measured in the scale. The higher the peak of the curve in the item information function, the more information an item provides. Items 15 and 16 in Factor 1 were the most informative, while Item 9 provided relatively less information than the other items.

#### Factor 2 IRT Validity Evidence

According to the item fit statistics, all the items' RMSEA values were less than 0.05. Based on this result, it was determined that the three items of Factor 2 provided an adequate fit according

**Table 3. Correlation level between initial and final administration of Troy Sleep Scale**

| Scale/ factors | n | r | p |
|---|---|---|---|
| Factor 1 Test* factor 1 retest administrations | 66 | 0.808 | <0.001 |
| Factor 2 Test* factor 1 retest administrations | 66 | 0.804 | <0.001 |
| Factor 3 Test* factor 1 retest administrations | 66 | 0.860 | <0.001 |
| Whole Scale Test* Whole Scale retest administrations | 66 | 0.878 | <0.001 |

**Table 4. Item parameters based on the generalized partial credit model (GPCM) for Factors 1, 2, and 3**

| **Factor 1** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Item no** | **a (SE)** | **b1 (SE)** | **b2 (SE)** | **b3 (SE)** | **b4 (SE)** | **S_$\chi^2$** | **df** | **RMSEA** |
| 9 | 1.475 (0.144) | -1.361 (0.130) | -0.189 (0.096) | 1.161 (1.120) | 2.767 (0.279) | 18.308 | 12 | 0.033 |
| 14 | 1.646 (0.168) | 0.014 (0.093) | 0.629 (0.103) | 1.491 (0.144) | 2.293 (0.241) | 17.196 | 13 | 0.026 |
| 15 | 4.856 (0.643) | -0.518 (0.064) | 0.471 (0.062) | 1.311 (0.084) | 2.361 (0.173) | 14.418 | 9 | 0.035 |
| 16 | 4.884 (0.656) | -0.681 (0.067) | 0.323 (0.061) | 1.114 (0.077) | 2.008 (0.131) | 14.823 | 7 | 0.048 |
| Iteration=22 | Log-likelihood: -2111.636 | p<0.05 | | | | | | |
| **Factor 2** | | | | | | | | |
| **Item no** | **a (SE)** | **b1 (SE)** | **b2 (SE)** | **b3 (SE)** | **b4 (SE)** | **S_$\chi^2$** | **df** | **RMSEA** |
| 21 | 2.809 (0.333) | -0.76 (0.079) | 0.214 (0.072) | 0.837 (0.082) | 1.897 (0.140) | 6.126 | 4 | 0.033 |
| 22 | 5.072 (0.846) | -0.499 (0.064) | 0.375 (0.061) | 1.214 (0.082) | 2.030 (0.137) | 8.243 | 4 | 0.046 |
| 23 | 2.165 (0.230) | -0.668 (0.086) | 0.156 (0.080) | 1.049 (0.100) | 1.893 (0.158) | 5.790 | 5 | 0.018 |
| Iteration=100 | Log-likelihood: -1711.795 | p<0.05 | | | | | | |
| **Factor 3** | | | | | | | | |
| **Item no** | **a (SE)** | **b1 (SE)** | **b2 (SE)** | **b3 (SE)** | **b4 (SE)** | **S_$\chi^2$** | **df** | **RMSEA** |
| 1 | 7.027 (1.968) | -1.479 (0.092) | -0.254 (0.058) | 0.534 (0.061) | 1.457 (0.089) | 14.594 | 11 | 0.026 |
| 2 | 2.527 (0.303) | -1.766 (0.127) | -0.448 (0.078) | 0.160 (0.075) | 1.401 (0.106) | 22.659 | 13 | 0.039 |
| 5 | 0.476 (0.062) | -1.694 (0.308) | 0.960 (0.281) | 1.275 (0.325) | 4.106 (0.695) | 31.338 | 24 | 0.025 |
| 13 | 1.197 (0.132) | 0.199 (0.118) | 0.829 (0.139) | 1.560 (0.197) | 1.695 (0.249) | 35.134 | 17 | 0.047 |
| Iteration=26 | Log-likelihood: -2417.189 | p<0.05 | | | | | | |

SE: Standart error, RMSEA: Root mean square error of approximation

to the generalizability theory (GPCM). The discrimination levels of Items 21 and 23 were optimal. The estimations based on the GPCM (log-likelihood ratio, $p<0.05$) confirmed the fit of the items. The item characteristic curves revealed that all options for items 21, 22, and 23 were functional. Upon analyzing the item information functions of Factor 2 items, it was found that Item 22 was the most informative. Item 23 provided relatively less information than the other items (Table 4).

**Factor 3 IRT Validity Evidence**

Furthermore, the RMSEA values for all the items in Factor 3 were also less than 0.05. Consequently, it was concluded that the four items of Factor 3 provided an adequate fit according to the GPCM. Items 2, 5, and 13 discrimination levels for Factor 3 were ideal. The GPCM (log-likelihood ratio, $p<0.05$) demonstrated the fit of the measurement tool items. The item characteristic curve of Item 13 revealed that the "most of the time" option was less effective than the other options. The remaining items functioned effectively with their respective options. Upon analyzing the item information functions of Factor 3 items, it was observed that Items 1, 2, and 5 were the most informative. Item 13 provided relatively less information than the other items (Table 4).

**Concurrent Validity**

The correlation between the scores obtained by the participants from the sleep assessment scale developed in this study and the scores obtained by the same participants from the PSQI was 0.73. In line with these results, the obtained correlation value corresponded to a robust and high-level correlation.[44-46] In this case, the concurrent validity level of the sleep assessment scale developed in this study was high. The Turkish form of the TSS is presented in Appendix 1.

## Discussion

In this study, the TSS scale was developed to assess subjective sleep problems and their consequences and its psychometric properties were analyzed. The results showed that this scale is valid and reliable for medical students in the Turkish population. The 5-point Likert scale consists of 11 items, and the construct validity analyses in the study showed that the scale had 3 factors. When the factors were analyzed, it was determined that the items of the first factor were related to "DSD" the items of the second factor were related to "ASD" and the items of the third factor was related to " (SQ; problems with initiating and maintaining sleep)". The CFA confirmed this structure. In addition, to evaluate its criterion-based validity, its correlation with the PSQI, accepted as the gold standard in the assessment of sleep quality, was examined, and a strong correlation was found. The reliability value was calculated for all scale items, and the test-retest reliability value was also high. To evaluate the structural validity, EFA indicated that the scale had three factors, and this three-factor structure explained 66% of the total variance. The KMO and Bartlett's sphericity test values were significant in EFA. The initial, subtracted, and corrected item-total values of the final 11 items in the TSS were between 0.45 and 0.79. These values are within acceptable limits when compared to other scales such as the PSQI, Athens Insomnia Scale, and Jenkins Sleep Scale.[47-49] Subsequent CFA supported this three-factor structure. The literature shows these fit indices confirm this structure.[27,28,30,32,34] Internal reliability results were similar to or better than the PSQI, Athens Insomnia Scale, and Jenkins Sleep Scale scores. When the test-retest correlation for the TSS was analyzed as another reliability indicator, a high correlation level was observed. Criterion-based validity analysis of the TSS was performed by examining its correlation with the PSQI scores, which are accepted as the gold standard for sleep. The results showed a strong correlation between the TSS and the PSQI. The present findings demonstrated the potential utility of a three-dimensional assessment of sleep disturbance by scoring three factors of the TSS rather than a single sleep quality index. Notably, the sleep efficiency, perceived sleep quality, and daily disturbance subscales of the PSQI's three-factor validation study were similar to those of the TSS.[50] The three-factor scale has the advantage of obtaining various assessments of sleep problems on a single scale. More information on the type and nature of sleep problems may be necessary to guide treatment choices. Despite these advantages, caution is advised when generalizing these findings, as they were obtained using only a young non-clinical population. It is also vital to recognize that these three factors may help to differentiate between those with and without sleep disorders. Sleep-related dysfunction, one of the factors on the TSS, is present in individuals with sleep disorders.[51] Sleep quality is closely related to an individual's physical, cognitive, or emotional functionality.[52,53] Sleep quality also affects cognitive areas, such as attention and memory (related to functionality).[54] In addition, quality of life, which is related to functionality, has been reported to be related to sleep quality.[55] There is a bidirectional relationship between mood and sleep disorders; sleep and symptoms are included in the diagnostic criteria for mood disorders, and sleep disorders are risk factors for the emergence and recurrence of depression.[56] Taken together, sleep disturbances are closely related to mood symptoms and functioning beyond sleep initiation, maintenance, and adequate sleep duration, and this has made this sleep scale helpful. The findings of this study should be evaluated considering its limitations. First, since it was conducted with university students and residents, it is not representative of the general population because it was administered to a limited age group. Second, this study was conducted on individuals with self-efficacy. This may limit the ability to determine the validity of these results in a broader population. Third, the scale was not administered to a clinical sample; it must be investigated for sleep disturbances due to mental disorders or other medical illnesses. Fourth, using PSQI as the only scale for concurrent validity is another weakness of the study. Despite these limitations, the study also has strengths. The fact that the scale was prepared considering Turkish culture and language features increased the comprehensibility of the scale. To our knowledge, a scale related to sleep in Turkish has not been developed before. In addition, the fact that IRT analysis was performed while developing the scale is another strength of this study.

## Conclusion

In conclusion, our research demonstrated that the TSS is a valid and dependable instrument for assessing sleep. This newly devised scale exhibited desirable properties, including self-administration, internal consistency, reliability, and construct validity, in a sample of university students. Consequently, it is a sound psychometric measurement tool for medical student populations. However, future studies are necessary to evaluate the scale's psychometric properties across various age groups, such as adolescents and the elderly, as well as in clinical samples with sleep disorders.

### Ethics

**Ethics Committee Approval:** The research was completed as follows. The approval of Çanakkale Onsekiz Mart University Clinical Research Ethics Committee (approval number: 2021-07, date: 20.10.2021) was obtained.
**Informed Consent:** Informant consent form was obtained from all participants, indicating their agreement to participate.

### Acknowledgments

We thank the students of Çanakkale Onsekiz Mart University Faculty of Medicine, who agreed to be included in the study.

### Footnotes

### Authorship Contributions

Concept: E.A., N.B., Design: E.A., N.B., Ç.T., Data Collection or Processing: E.A., N.B., Analysis or Interpretation: E.A., Ş.A.K., Ç.T., Literature Search: E.A., N.B., Ş.A.K., Ç.T., Writing: E.A., N.B., Ş.A.K., Ç.T.

**Conflict of Interest:** No conflict of interest was declared by the authors.
**Financial Disclosure:** The authors declared that this study received no financial support.

## References

1. K Pavlova M, Latreille V. Sleep disorders. Am J Med. 2019;132(3):292-299.
2. Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Med Rev. 2002;6(2):97-111.
3. Roth T. Insomnia: definition, prevalence, etiology, and consequences. J Clin Sleep Med. 2007;3(5 Suppl):S7-10.
4. American Academy of Sleep Medicine. The international classification of sleep disorders. 3rd ed. (ICSD-3). Darien IL, American Academy of Sleep Medicine; 2014.
5. Yetkin S. Uykusuzluk (İnsomni). Uyku Bozuklukları Tanı ve Tedavi Kitabı. In: Akıncı E, Orhan FÖ, Demet MM (editors). 1st ed. Ankara: Türkiye Psikiyatri Derneği Yayınları; 2016. pp. 67-86.
6. Jiang XL, Zheng XY, Yang J, et al. A systematic review of studies on the prevalence of insomnia in university students. Public Health. 2015;129(12):1579-1584.
7. Russell K, Allan S, Beattie L, Bohan J, MacMahon K, Rasmussen S. Sleep problem, suicide and self-harm in university students: A systematic review. Sleep Med Rev. 2019;44:58-69.
8. Carrión-Pantoja S, Prados G, Chouchou F, et al. Insomnia symptoms, sleep hygiene, mental health, and academic performance in Spanish University students: a cross-sectional study. J Clin Med. 2022;11(7):1989.
9. Mbous YPV, Nili M, Mohamed R, Dwibedi N. Psychosocial correlates of insomnia among college students. Prev Chronic Dis. 2022;19:E60.
10. Karakoç AGDFY, Dönmez PDL. Ölçek geliştirme çalışmalarında temel ilkeler. TED. 2014;13(40):39-34.
11. Conley S, Knies A, Batten J, et al. Agreement between actigraphic and polysomnographic measures of sleep in adults with and without chronic conditions: A systematic review and meta-analysis. Sleep Med Rev. 2019;46:151-160.
12. Yi H, Shin K, Shin C. Development of the sleep quality scale. J Sleep Res. 2006;15(3):309-316.
13. Bozkurt A, Uygur ÖF, Demiröğen EY, Akıncı MA. A review of Turkish self-reporting ınstruments for sleep problems in children, adolescents, and adults. Eurasian J Med. 2023;55(1):S55-S60.
14. Ağargün MY, Kara H, Anlar Ö. Pittsburgh Uyku Kalitesi İndeksi'nin geçerliği ve güvenirliği. Türk Psikiyatri Derg. 1996;7(2):107-115.
15. Boysan M, Güleç M, Beşiroğlu L, Kalafat T. Psychometric properties of the insomnia severity index in Turkish sample. Alpha Psychiatry. 2010;11(3):248-252.
16. Uygur H, Tekdemir R, Uygur OF, et al. Psychometric properties of the Turkish reduced morningness and eveningness questionnaire. Chronobiol Int. 2024;41(5):632-646.
17. Uygur H, Ahmed O, Uygur OF, et al. Validity and reliability of the Turkish version of the sleep condition indicator: a clinical screening ınstrument based on the DSM-5 criteria for insomnia. Nat Sci Sleep. 2024;16:63-74.
18. Uygur OF, Uygur H, Chung S, et al. Validity and reliability of the Turkish version of the Glasgow Sleep Effort Scale. Sleep Med. 2022;98:144-151.
19. Uygur ÖF, Orhan FÖ, Uygur H, Kandeger A, Hursitoglu O. Psychometric properties of the Turkish version of the Anxiety and Preoccupation about Sleep Questionnaire in clinical and non-clinical samples. Sleep Sci. 2022;15(1):68-74.
20. Hair JF, Black WC, Babin BJ, et al. Multivariate data analysis. Harlow: Pearson Education Limited; 2014.
21. Aybek EC, Toraman C. How many response categories are sufficient for Likert type scales? an empirical study based on the item response theory. International Journal of Assessment Tools in Education. 2022;9(2):534-547.
22. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193-213.
23. Korkmaz S, Goksuluk D, Zararsiz G. MVN: An r package for assessing multivariate normality. The R Journal. 2014;6(2):151-162.
24. Kaiser HF, Rice J. Little jiffy, mark IV. Educ Psychol Meas. 1974;34:111-117.
25. Warner RM. Applied statistics, from bivariate through multivariate techniques. California, SAGE Publications, Inc.; 2013.
26. Nunnally JC, Bernstein IH. Psychometric theory. 3rd. ed. New York: McGraw-Hill; 1994.
27. Anderson JC, Gerbing DW. The effect of sampling error on convergence, improper solutions, and goodness of fit ındices for maximum likelihood confirmatory factor analysis. Psychometrika. 1984;49:155-73.
28. Bentler PM. Comparative fit indexes in structural models. Psychol Bull. 1990;107(2):238-246.
29. Hooper D, Coughlan J, Mullen MR. Structural equation modelling: Guidelines for determining model fit. Electron J Bus Res Methods. 2008;6(1):53-60.
30. Hu LT, Bentler PM. Cut off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling. 2009;(6)1:1-55.
31. Kline TJB. Psychological testing, a practical approach to design and evaluation. The USA: Sage; 2005.

32. Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indices in confirmatory factor analysis: the effect of sample size. Psychol Bull. 1998;103(3):391-410.
33. Tabachnick BG, Fidell LS. Using multivariate statistics. Pearson Education;2013.
34. Vieira AL. Preparation of the analysis. Interactive LISREL in practice. 1st ed. London: Springer; 2011.
35. Embretson SE, Reise SP Item response theory for psychologists. 1st ed. New York: Psychological Press; 2000.
36. Zhao Y. Approaches for addressing the fit of item response theory models to educational test data. PhD dissertation, University of Massachusetts Amherst, Holland; 2008. Available from: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://core.ac.uk/download/pdf/334980854.pdf
37. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of item response theory. 1. st ed. London: Sage Publication; 1991.
38. Yen WM. Scaling performance assessments: Strategies for managing local item dependence. J Educ Meas. 1993;30(3):187-213.
39. Chalmers RP. mirt: A multidimensional item response theory package for the R environment. J Stat Softw. 2012;48(6):1-29.
40. Flannery WP, Reise SP, Widaman KF. An item response theory analysis of the general and academic scales of the self-description questionnaire II. J Res Pers. 1995;29(2):168-188.
41. Hambleton RK. Guidelines for adapting educational and psychological test: a progress report. Eur J Psychol Assess. 1994;10(3):229-244.
42. Stout WF. A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. Psychometrika. 1990;55:293-325.
43. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res. 2007;16:5-18.
44. Pallant J. SPSS Survival Manual. 6th ed. USA: McGraw-Hill Education; 2016.
45. Schober P, Boer C, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. Anaesth Analg. 2018;126(5):1763-1768.
46. Rumsey DJ. Statistics all-in-one for dummies. New Jersey: John Wiley & Sons, Inc.; 2022.
47. Carpenter JS, Andrykowski MA. Psychometric evaluation of the Pittsburgh Sleep Quality Index. J Psychosom Res. 1998;45(1):5-13.
48. Jenkins CD, Stanton BA, Niemcryk SJ, Rose RM. A scale for the estimation of sleep problems in clinical research. J Clin Epidemiol. 1988;41:313-321.
49. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. J Psychosom Res. 2000;48(6):555-560.
50. Cole JC, Motivala SJ, Buysse DJ, Oxman MN, Levin MJ, Irwin MR. Validation of a 3-factor scoring model for the Pittsburgh Sleep Quality Index in older adults. Sleep. 2006;29(1):112-116.
51. Valenza MC, Cabrera-Martos I, Martín-Martín L, Pérez-Garzón VM, Velarde C, Valenza-Demet G. Nursing homes: impact of sleep disturbances on functionality. Arch Gerontol Geriatr. 2013;56(3):432-436.
52. Kim M, Yoshida H, Sasai H, Kojima N, Kim H. Association between objectively measured sleep quality and physical function among community-dwelling oldest old Japanese: A cross-sectional study. Geriatr Gerontol Int. 2015;15(8):1040-1048.
53. Liao H, Liao S, Gao YJ, Mu JP, Wang X, Chen DS. Correlation between Sleep Time, Sleep Quality, and emotional and cognitive function in the elderly. Biomed Res Int. 2022;2022:9709536.
54. Siddarth P, Thana-Udom K, Ojha R, et al. Sleep quality, neurocognitive performance, and memory self-appraisal in middle-aged and older adults with memory complaints. Int Psychogeriatr. 2021;33(7):703-713.
55. Tel H. The quality of sleep and the quality of life among the elderly people. Eur Psychiatry. 2011;26(S2):405-405.
56. Riemann D; Workshop Participants. Does effective management of sleep disorders reduce depressive symptoms and the risk of depression? Drugs. 2009;69 (Suppl 2):43-64.

**Appendix 1. English Forms of Troy Sleep Scale**

**The Troy Sleep Scale (English form)***

**Answer the following questions taking into account your sleep habits in the last 1 (ONE) month:**
**1: Never**
**2: Rarely**
**3: Sometimes**
**4: Most of the time**
**5: Always**

| Item no. | Items | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|---|
| 1 | The sleep problem I experience at night prevents me from doing my work during the day. | | | | | |
| 2 | My night-time sleep problems impair my daytime functioning. | | | | | |
| 3 | I cannot focus on my work during the day due to insomnia at night. | | | | | |
| 4 | I suffer from forgetfulness during the day due to sleep problems at night. | | | | | |
| 5 | I am constantly nervous and irritable due to sleep problems. | | | | | |
| 6 | My tolerance for people is decreasing due to sleep problems. | | | | | |
| 7 | I'm easily depressed because of the sleep problems I've been having. | | | | | |
| 8 | I have trouble falling asleep at night. | | | | | |
| 9 | I spend much time in bed at night until I fall asleep. | | | | | |
| 10 | I have anxiety about not being able to sleep when I go to bed. | | | | | |
| 11 | If I wake up at night, I find it difficult to fall asleep again. | | | | | |

*The Troy Sleep Scale was developed in Turkish. English translations of the scale items are given for international readers to understand, but they should not be considered an English cultural adaptation