

RESEARCH

Open Access



# The RU\_SATED as a measure of sleep health: cross-cultural adaptation and validation in Chinese healthcare students

Runtang Meng<sup>1,2\*</sup>, Lu Dong<sup>3</sup>, Joseph M. Dzierzewski<sup>4,5</sup>, Stefanos Mastrotheodoros<sup>6,7</sup>, Menglu Cao<sup>8,9</sup>, Bilin Yu<sup>10</sup>, Jue Wang<sup>11</sup>, Boxiong Gong<sup>12</sup>, Jingjing Li<sup>13</sup> and Karen Spruyt<sup>14</sup>

## Abstract

**Background** The RU\_SATED scale is a multidimensional instrument measuring sleep health, consisting of Regularity, Satisfaction, Alertness, Timing, Efficiency, Duration dimensions. We adapted and validated the Chinese RU\_SATED (RU\_SATED-C) scale.

**Methods** The RU\_SATED-C scale was developed through a formal linguistic validation process and was validated in an observational longitudinal survey design. Healthcare students completed the RU\_SATED scale, Sleep Quality Questionnaire, and Patient Health Questionnaire-4 among two sites of Hangzhou and Ningbo, China. Psychometric assessments included structural validity, longitudinal measurement invariance, convergent and divergent validity, internal consistency, and test–retest reliability.

**Results** A total of 911 healthcare students completed the RU\_SATED-C scale at baseline (Time 1, T1) and follow-up (Time 2, T2) with an average time interval of 7 days + 5.37 h. Confirmatory factor analysis (CFA) confirmed a single-factor model and resulted in an acceptable model fit. The two-factor model previously found in the Japanese version fit better than the one-factor model, whereas the one-factor model fit had a better fit than the two-factor model found in the English version. Longitudinal CFA resulted in negligible changes in fit indices for four forms of increasingly restrictive models and supported that a single-factor model was equivalent over time. The data also endorsed longitudinal measurement invariance among the two-factor models found in the English and Japanese samples. The RU\_SATED-C scale total score displayed a moderately strong negative correlation with sleep quality; however, negligible associations were observed with anxiety and depression. Ordinal Cronbach's alpha and Ordinal McDonald's omega at T1 and T2 ranged from suboptimal to acceptable. The RU\_SATED-C scale and all items were significantly correlated across time intervals.

**Conclusion** The RU\_SATED-C scale is an easy-to-use instrument with potentially valid data for the measurement of multidimensional sleep health. Use of the RU\_SATED-C scale can help raise awareness of sleep health and could pave the way for important efforts to promote healthy sleep.

**Keywords** RU\_SATED, Sleep health, Cross-cultural adaptation, Validation, Psychometrics, Longitudinal

\*Correspondence:

Runtang Meng  
mengruntang@hznu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

Better sleep is a cornerstone of better health. To date, sleep health is recognized as a major global public health concern; thus, improving sleep health is a necessary step toward achieving better health [1]. While significant resources have been invested in individual and population-level interventions to remedy unhealthy lifestyle factors such as nutrition, exercise, and smoking control, programs concentrating on sleep health have been notably scarce [2]. Sleep health has been defined as “a multidimensional pattern of sleep-wakefulness, adapted to individual, social, and environmental demands, that promotes physical and mental well-being” [3]. The sleep health framework was developed based on an extensive review of the scientific literature, including a review of specific dimensions of sleep and their association with numerous health outcomes, providing a comprehensive framework for examining sleep health. This multidimensional sleep health framework differs from traditional operationalizations of sleep in medicine in that it does not focus on identifying and treating sleep disorders; instead, sleep health promotes a positive framework that views sleep as a multidimensional construct considering positive attributes of sleep along important dimensions—sleep duration, sleep continuity or efficiency, timing, alertness/sleepiness, satisfaction/quality, and regularity—that are associated with physical and mental health. Alternatively, sleep health is broadly defined as a pattern of sleep that is associated with optimal physical and mental health, and is not merely the absence of a sleep disorder, encompassing sleep duration and quality in non-disordered sleepers [4].

Assessing and promoting multidimensional sleep health may offer potential benefits [2, 3, 5–9]. First, consistent with the World Health Organization’s (WHO) definition of health, the concept of sleep health broadens key dimensions of good sleep and enables individuals to comprehensively quantify and modify the level of sleep health. Second, the conceptual framework of sleep health provides useful building blocks and frameworks that facilitate developing new sleep health instruments, as a foundation for adding additional domains. Third, the sleep health framework avoids simply dichotomizing the sleep conditions of individuals into healthy and unhealthy by capturing graduations in sleep, noting that the sleep health of individuals exists on a continuum. Finally, identifying and measuring sleep health instead of only assessing and treating sleep disorders may increase awareness, optimize personalized sleep recommendations, and enhance evidence-based self-management of sleep behaviors, more significantly, allowing for earlier interventions to prevent the adverse downstream effects of suboptimal sleep.

Sleep can be assessed via objective and subjective measures including self-report questionnaires or sleep diaries, actigraphy, and home or laboratory polysomnography [3]. While there have been numerous instruments that assess sleep disturbance or sleep quality in clinical and research settings, instruments that measure sleep health are rare. A short, practical self-report instrument for the measurement of sleep health called the SATED (Satisfaction, Alertness, Timing, Efficiency, Duration) (v1.0) scale was proposed in 2014 [3], with subsequent expansion to include an additional dimension/item assessing Regularity. The current instrument is now called the RU\_SATED (v2.0) scale [3]. Another sleep health assessment instrument was developed by the National Sleep Foundation—Sleep Health Index (SHI) [10]. The SHI is a 12-item instrument designed to capture three discrete dimensions of sleep health: duration, quality, and disorders. Noteworthy, the RU\_SATED scale has a richer conception of sleep health, and is half the length of the SHI, while including more theoretical dimensions of sleep health. These six dimensions of the RU\_SATED scale are appropriate indicators of sleep health for several reasons. First, each has been associated with important health outcomes, albeit with somewhat different outcomes for each dimension. Second, they can each be expressed in positive terms, i.e., we can characterize their “better” directions. This is not to say that these dimensions are all unidirectional. It is also important to acknowledge that, while these dimensions can be expressed in positive terms, the supporting studies largely focus on their negative directions and consequences; there have been few studies specifically examining the potential benefits of good sleep.

Poor sleep is common among healthcare students, with prevalence estimates suggesting higher rates of poor sleep than in non-healthcare students and the general population [11–14]. A 2022 meta-analysis reported that the prevalence of sleep problems among healthcare students was close to thirty percent in China [15]. The domains of sleep health that are typically poor in healthcare students due to academic overload and rigorous training are: insufficient sleep duration, poor sleep quality, and excessive daytime sleepiness amongst others [11, 13, 14]. The dire situation for healthcare students requires urgent attention and effective intervention, such as regular monitoring and screening of poor sleep health. Healthcare students, therefore, comprise an important and interesting population in which to test the RU\_SATED scale.

To date, the RU\_SATED scale has been cross-culturally adapted and validated in at least five languages: Portuguese (2018) [16], English (2019) [9], Spanish (2020) [17], French (2021) [18], and Japanese (2022) [19]. In the current study, we developed the Chinese version of the

RU\_SATED (RU\_SATED-C) scale and conducted a longitudinal observational design in a sample of healthcare students in China. The primary study aims were to (i) develop a Chinese version of the RU\_SATED scale and (ii) assess the main measurement properties of the RU\_SATED-C scale: structural validity, longitudinal measurement invariance, convergent and divergent validity, internal consistency, and test–retest reliability.

## Methods

### Linguistic validation of the Chinese RU\_SATED (RU\_SATED-C) scale

Using the formal procedure for linguistic validation, the original RU\_SATED (v2.0) scale was translated into Chinese following Mapi instructions [20], including translation by two separate translators, qualitative interviews to determine people’s understanding of the questions in the new language (i.e., Chinese), and back-translation by two other translators. The linguistic validation process is essential to ensure that the RU\_SATED (v2.0) scale is actually measuring what it was intended to measure in the newly translated language.

*Step 1* Preparation: Initial planning and actions carried out before the translation process began included conceptual analysis of the original questionnaire and application for approval to use the original questionnaire. After obtaining permission from the original author (Prof. Daniel J. Buysse, DJB) of the RU\_SATED scale, an e-contract was signed with the University of Pittsburgh for the preparation of the Chinese version of the RU\_SATED scale.

*Step 2* Forward translation: The original RU\_SATED scale was translated into Chinese independently by two Chinese native speakers, a psychologist (co-author, MC), and a linguist (BY) with a high level of fluency in both English and Chinese. A panel of five local clinical and research experts (MC, BY, JW, BG, and RM) checked and compared the two translations to create the preliminary initial translated form of the scale.

*Step 3* Backward translation: The back-translation into English was undertaken by two independent highly proficient bilingual English-Chinese speakers (i.e., a behavioral scientist and clinical psychologist [LD] and a behavioral scientist and physician [JL]), and was made independently of the forward translation. The original author (DJB) reviewed the two back-translations, which were rated as satisfactory.

*Step 4* Pilot Testing: Eight Chinese healthcare students were surveyed to see whether they could understand the meaning of the translated items, instrument instructions, and answer choices. Pilot testing revealed that no explanations were required, with all eight individuals confirming full understanding of the RU\_SATED-C scale.

*Step 5* Proofreading and finalization: The research team (RM, LD, JL, MC, BY, JW, BG, and DJB) involved in the forward translation, consolidation, and backward translation processes evaluated the pre-final version of the scale and confirmed the equivalence between the Chinese and English versions. The final Chinese RU\_SATED scale was delivered to the original author (DJB) and is housed electronically at the University of Pittsburgh.

### Participants and procedures

For this validation study, routinely collected data were available from two sample sites (Hangzhou and Ningbo, China) and contained an assessment of sleep using the below three measures from December 2020 until January 2021. The trained investigators were responsible for the conduct of the survey and its onsite quality control. Self-administered paper-and-pencil survey was centralized at recess or evening self-study. Healthcare students were recruited by applying a stratified random sampling approach based on their academic years and majors [21]. Inclusion criteria: individuals who were able to read simplified Chinese and communicate in Mandarin. Exclusion criteria: 1) people who were reluctant to participate; 2) those who had difficulty understanding the study procedures. Given that a retest interval of two to 14 days is usually adequate [22, 23] and reproducibility of health status measures intended for longitudinal use may best be measured at intervals of 1–2 weeks [24]. 976 healthcare students responded to the baseline assessment (Time 1, T1) and 951 completed a follow-up assessment approximately 7 days later (Time 2, T2). A total of 911 questionnaires were matched by student ID at two time points. Each participant received 2 CNY (around 0.30 US dollars) upon completion of the study as compensation for their time.

### Measures

#### RU\_SATED scale

Sleep health was assessed using the RU\_SATED (Regularity, Satisfaction, Alertness, Timing, Efficiency, Duration) scale, consisting of six key dimensions of sleep health that are consistently associated with various health outcomes [3]. The scale consists of six items/dimensions of sleep health and queries about sleep during the previous month. Each item is scored from 0 to 2 on a three-point Likert scale, with 0 for “never” or “rarely,” 1 for “sometimes,” and 2 for “usually” or “always.” Scoring entails summing the scores of the individual items, with total scores ranging from 0 (poor sleep health) to 12 (good sleep health).

### **Sleep quality questionnaire (SQQ)**

Sleep quality was measured by the Sleep Quality Questionnaire (SQQ) [25]. This questionnaire evaluates two components—daytime sleepiness (four items) and sleep difficulty (six items)—of sleep quality in the last month. Each item is scored from 0 (strongly disagree) to 4 (strongly agree) on a five-point Likert scale. The overall SQQ score ranges from 0 to 40, with higher scores indicating poorer sleep quality. Psychometric data for the Chinese version of the Sleep Quality Questionnaire (SQQ-C) reveal adequate measurement properties in multi-site studies [21, 26–28].

### **Patient health questionnaire-4 (PHQ-4)**

A self-report version of the Primary Care Evaluation of Mental Disorders (PRIME-MD) called the Patient Health Questionnaire (PHQ) was developed and validated in two large studies for use with general adult samples [29]. The PHQ-4 is a validated measure of mental health symptoms consisting of the first two items of the PHQ-9 and the GAD-7, respectively [30]. Each item is scored from 0 (not at all) to 3 (nearly every day). The total score ranges from 0 to 6, with a higher score indicating greater severity of anxiety or depression over the last two weeks. The Chinese version of the PHQ-4 (PHQ-4-C) and its instruction manual are publicly available and no permission is required for use [31].

## **Statistical analysis**

### **Data preparation**

Data were checked for data entry errors, missing data, or the presence of extreme outliers. Frequencies (%) were calculated for categorical variables, whereas means and standard deviations were computed for continuous variables. Multivariate normality was assessed via skewness and kurtosis. Data analyses were performed with JASP (v.0.16.1) and R (v.4.1.2). The packages “*naniar* v 1.0.0” [32], “*MVN* v.5.9” [33], “*lavaan* v.0.6-9” [34], “*semTools* v.0.5-5” [35], “*irr* v.0.84.1” [36], and “*ufs* v.0.5.2” [37] under RStudio were utilized to conduct the missing value analysis, multivariate normality tests, confirmatory factor analysis (CFA), longitudinal CFA (LCFA), intraclass correlation coefficient (ICC), and Cronbach’s alpha as well as McDonald’s omega. After missing value analysis, of the 911 participants, 898 (98.6%) had no missing data, while 13 (1.43%) had some missing data. Of the total 12 RU\_SATED-C scale items (T1 and T2) missingness ranged from 0.11% to 0.44%. Missingness was therefore considered negligible, and listwise deletion was applied for factor analysis (i.e., structural validity and longitudinal measurement invariance) and reliability analysis (i.e., internal consistency and test–retest reliability). In other analyses (N=911), convergent and divergent validity and

reliability for other measures, missing data was replaced by the mean or median of observed values given that missing data rates did not exceed 10% [38, 39] or 5% [40]. We assessed the below measurement properties of the measures, adhering to the COnsensus-based Standards for the selection of health Measurement INSTRUMENTS (COSMIN) taxonomy and guideline [41, 42].

### **Structural validity**

Structural validity measures the degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct measured [42]. The structural validity of the RU\_SATED-C scale was assessed by CFAs. Because the six items are supposed to measure one construct (sleep health), we expected that all items would load on a single factor [3], similar to that of findings in the Portuguese, Spanish, and French samples [16–18]. The single-factor structure of the RU\_SATED-C scale was evaluated across two points in time independently (T1 and T2; a cross-sectional CFA at each time point), as well as through a LCFA approach. We applied the mean and variance adjusted diagonally weighted least squares (DWLS) estimator based on the polychoric correlation matrix to examine unidimensionality, given those responses to items in the RU\_SATED-C scale are ordinal [43, 44]. In addition to the one-factor model, we examined the fit of the two-factor models that were found in the English and Japanese samples [9, 19].

Model fit indices include the chi-squared test statistic and its associated degrees-of-freedom (*df*) and *p*-value [40]. However, considering that the chi-squared test is known to be very sensitive to large sample sizes, we also included additional relevant fit indices: comparative fit index (CFI), Tucker–Lewis index (TLI), root means square error of approximation (RMSEA) and its corresponding 90% confidence interval. Scaled fit indices instead of unscaled indices were reported in this paper because the former is considered more precise [45]. Following the recommended guidelines, we considered acceptable model fit if  $CFI \geq 0.90$ ,  $TLI \geq 0.90$ , and  $RMSEA \leq 0.08$  [40, 46]; good model fit if  $CFI \geq 0.95$  or  $TLI \geq 0.95$ , and  $RMSEA \leq 0.06$  [41, 47].

### **Longitudinal measurement invariance**

Following confirmation of the single-factor and two-factor structure of the RU\_SATED-C scale, we explored longitudinal measurement invariance (LMI) in the matched sample (N=898) across time. LCFA was used to examine four forms of increasingly restrictive invariance: configural invariance (same pattern of free loadings), metric or weak invariance (common loadings over time), scalar or strong invariance (common loadings and intercepts over time), and strict or residual invariance (common

loadings, intercepts, and residual variances over time). The fit of two nested models can be compared by taking the difference of the fit indices. However, the scaled chi-square difference suffers from the same issues of significance with large sample sizes as the minimum fit function statistic [48]. Hence, we focused on changes in model fit according to CFI, TLI, and RMSEA when the scaled chi-square difference was significant [48]. Following the recommended cut-off criteria, we considered an acceptable model fit for more restrictive invariant models in the following circumstances:  $\Delta\text{CFI} \leq 0.010$ ,  $\Delta\text{TLI} \leq 0.010$ , and  $\Delta\text{RMSEA} \leq 0.015$  [49]. If at least two out of three changes in fit indices meet the cut-off criteria, we considered that longitudinal measurement invariance held [50].

### Convergent and divergent validity

For assessing convergent and divergent validity, we hypothesized that the RU\_SATED-C scale total score would have a moderately strong negative correlation ( $-0.50 < r < -0.30$ , Spearman) with the SQQ-C, given that both instruments measure sleep-related constructs, and a weak negative correlation ( $-0.30 < r < 0$ , Spearman) with the PHQ-4-C, due to the theoretically distinct nature of sleep and mental health constructs [41].

### Internal consistency

Internal consistency measures the degree of interrelatedness among measure items [42]. Internal consistency of the RU\_SATED-C scale was determined by calculating ordinal Cronbach's alpha and McDonald's omega to accommodate categorical data [51]. Values greater or equal to 0.70 was considered sufficient evidence for internal consistency [52].

### Test-retest reliability

Test-retest reliability reflects the consistency in measurement taken by the same instrument, on the same subjects, under the same or very similar conditions [53]. ICC estimated by a two-way mixed model was used to evaluate test-retest reliability of the RU\_SATED-C scale. An  $\text{ICC} < 0.40$  was considered poor,  $0.40 \leq \text{ICC} < 0.60$  fair,  $0.60 \leq \text{ICC} < 0.75$  good, and  $\text{ICC} \geq 0.75$  excellent [54].

## Results

### Descriptive

We analyzed final data from the matched sample of 911 healthcare students (complete data response rate 93.34% and 95.79% for T1 and T2, respectively). The average time interval between T1 and T2 was 7 days + 5.37 h. The mean age of participants was  $19.66 \pm 1.45$  years, ranging from 17 to 31 years. Additional descriptive information is presented in Table 1. Means, standard deviations,

**Table 1** Descriptive characteristics of study cohort (N=911)

Characteristics	n (%)
<i>Gender</i>	
Female	736 (80.790)
Male	175 (19.210)
<i>Age</i>	
≤ 18	200 (21.954)
19	217 (23.820)
20	315 (34.577)
21	118 (12.953)
≥ 22	61 (6.696)
<i>Major</i>	
Clinical medicine	121 (13.282)
Preventive medicine	96 (10.538)
Nursing	270 (29.638)
Pharmacy	94 (10.318)
Health policy and management	139 (15.258)
Health services and management	81 (8.891)
Midwifery	69 (7.574)
Stomatology	41 (4.501)
<i>Academic stage</i>	
Junior college students	316 (34.687)
Undergraduate	554 (60.812)
Postgraduate	41 (4.501)

skewness, kurtosis, and amount of missing data at T1 and T2 are presented in Additional file 1: Table S1.

### Structural validity and longitudinal measurement invariance

The CFA at each time point resulted in an acceptable fit for the single-factor model [CFI=0.958, TLI=0.929, RMSEA=0.054 (0.035, 0.075) and CFI=0.967, TLI=0.945, RMSEA=0.058 (0.039, 0.079)] at T1 and T2, respectively (Table 2). The fit indices were CFI=0.976, TLI=0.956, RMSEA=0.043 (0.021, 0.066) and CFI=0.974, TLI=0.952, RMSEA=0.055 (0.035, 0.077)] at T1 and T2, respectively, indicating good fit of the two-factor model found in the Japanese samples. Similarly, the fit indices were CFI=0.957, TLI=0.919, RMSEA=0.058 (0.038, 0.080) and CFI=0.966, TLI=0.937, RMSEA=0.063 (0.043, 0.084)] at T1 and T2, respectively, indicating acceptable fit of the two-factor model found in the English samples. After the scaled chi-squared difference test, we compared the one-factor and two-factor models fit at the same time point. The two-factor model in the Japanese version fit outperformed that of the one-factor model and the models differed substantially, and the one-factor model fit was superior to that of the two-factor model in the English version and yet the model difference was negligible.

**Table 2** Fit indices for alternative models of the RU\_SATED-C at T1 and T2 (N = 898)

Model	$\chi^2$ (df)	P	CFI	TLI	RMSEA (90% CI)	Scaled Chi-squared difference test	
						$\Delta\chi^2$ ( $\Delta df$ )	P
One-factor (T1)	32.974 (9)	< 0.001	0.958	0.929	0.054 (0.035, 0.075)		
One-factor (T2)	36.421 (9)	< 0.001	0.967	0.945	0.058 (0.039, 0.079)		
Two-factor (T1, Japanese)	21.317 (8)	< 0.001	0.976	0.956	0.043 (0.021, 0.066)	13.399 (1)	< 0.001
Two-factor (T2, Japanese)	29.623 (8)	< 0.001	0.974	0.952	0.055 (0.035, 0.077)	7.579 (1)	0.006
Two-factor (T1, English)	32.407 (8)	< 0.001	0.957	0.919	0.058 (0.038, 0.080)	0.656 (1)	0.418
Two-factor (T2, English)	36.313 (8)	< 0.001	0.966	0.937	0.063 (0.043, 0.084)	0.155 (1)	0.694
Threshold	N/A	> 0.05	> 0.900	> 0.900	< 0.080		

Report "Used/Standard" fit indices due to use of listwise deletion; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation; CI, confidence interval; T1, Time 1; T2, Time 2; N/A, N/A, not applicable

The LCFA provided strong evidence for invariance (Table 3). Configural invariance was supported by fit indices meeting requirements for excellent model fit [CFI = 0.987, TLI = 0.982, RMSEA = 0.047 (0.038, 0.056)] for the single-factor model, [CFI = 0.991, TLI = 0.986, RMSEA = 0.041 (0.032, 0.051)] for the two-factor model in the Japanese version, [CFI = 0.987, TLI = 0.980, RMSEA = 0.051 (0.041, 0.060)] for the two-factor model in the English version, respectively. Successively stricter constraints on factor loadings (metric), loadings and intercepts (scalar), and loadings, intercepts, and residual variances (strict) revealed that all three invariance

models (metric, scalar, and strict measurement invariance) were supported with negligible changes in fit indices across time (all  $\Delta CFI \leq 0.010$ ,  $\Delta TLI \leq 0.010$ , and  $\Delta RMSEA \leq 0.015$ ). Hence, all four forms of longitudinal measurement invariance among the single-factor model and the two-factor models in Japanese and English versions were supported.

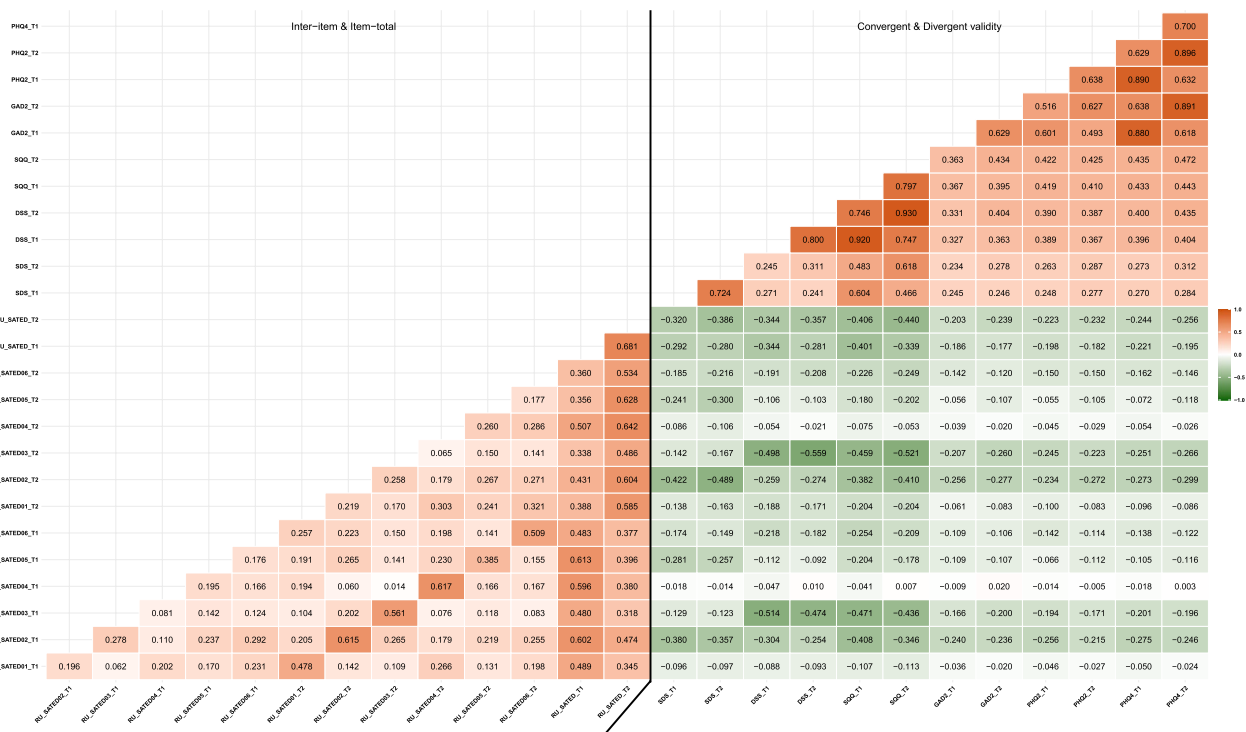
**Convergent and divergent validity**

The total scores for the RU\_SATED-C scale at T1 and T2 were  $8.286 \pm 2.148$  and  $8.375 \pm 2.230$ , respectively. The total scores for the SQQ-C at T1 and T2

**Table 3** Longitudinal measurement invariance of the RU\_SATED-C across time (N = 898)

Hypothesis	$\chi^2$ (df)	P	Scaled Chi-squared difference test		CFI	$\Delta CFI$	TLI	$\Delta TLI$	RMSEA (90% CI)	$\Delta RMSEA$
			$\Delta\chi^2$ ( $\Delta df$ )	P						
<i>One-factor</i>										
Configural	139.541 (47)	< 0.001			0.987		0.982		0.047 (0.038, 0.056)	
Metric	143.861 (52)	< 0.001	4.899 (5)	0.428	0.988	0.001	0.984	0.002	0.044 (0.044, 0.053)	- 0.003
Scalar	165.746 (57)	< 0.001	31.812 (5)	< 0.001	0.985	- 0.003	0.983	- 0.001	0.046 (0.038, 0.054)	0.002
Strict	185.003 (63)	< 0.001	25.691 (6)	< 0.001	0.983	- 0.002	0.983	0.000	0.046 (0.039, 0.054)	0.000
<i>Two-factor (Japanese)</i>										
Configural	106.216 (42)	< 0.001			0.991		0.986		0.041 (0.032, 0.051)	
Metric	110.776 (46)	< 0.001	5.416 (4)	0.247	0.991	0.000	0.987	0.001	0.040 (0.030, 0.049)	- 0.001
Scalar	123.231 (50)	< 0.001	18.179 (4)	0.001	0.990	- 0.001	0.987	0.000	0.040 (0.031, 0.049)	0.000
Strict	146.819 (56)	< 0.001	33.438 (6)	< 0.001	0.988	- 0.002	0.985	- 0.002	0.043 (0.034, 0.051)	0.003
<i>Two-factor (English)</i>										
Configural	138.104 (42)	< 0.001			0.987		0.980		0.051 (0.041, 0.060)	
Metric	138.546 (46)	< 0.001	0.535 (4)	0.970	0.987	0.000	0.982	0.002	0.047 (0.038, 0.057)	- 0.004
Scalar	154.578 (50)	< 0.001	23.596 (4)	< 0.001	0.986	- 0.001	0.981	- 0.001	0.048 (0.040, 0.057)	0.001
Strict	164.531 (56)	< 0.001	14.025 (6)	0.029	0.985	- 0.001	0.983	0.002	0.046 (0.038, 0.055)	- 0.002
Threshold	N/A	> 0.05	N/A	> 0.05	> 0.90	< 0.010	> 0.90	< 0.010	< 0.08	< 0.015

Report "Used/Standard" fit indices due to use of listwise deletion; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation; CI, confidence interval; N/A, not applicable



**Fig. 1** Inter-item and item-total, convergent and divergent correlations between the RU\_SATED-C scale, SQQ-C and PHQ-4-C (N = 911). *Note:* Spearman correlations; T1, Time 1; T2, Time 2; RU\_SATED, Regularity, Satisfaction, Alertness, Timing, Efficiency, Duration; SQQ, Sleep Quality Questionnaire; DSS, Daytime Sleepiness Subscale; SDS, Sleep Difficulty Subscale; PHQ-4, Patient Health Questionnaire-4; GAD-2, Generalized Anxiety Disorder-2; PHQ-2, Patient Health Questionnaire-2

were  $18.058 \pm 6.265$  and  $17.903 \pm 6.343$ , respectively. The total scores for the PHQ-4-C at T1 and T2 were  $3.501 \pm 2.214$  and  $3.318 \pm 2.016$ , respectively. The Spearman correlations of the total SQQ-C scores with the total RU\_SATED-C scale scores were  $-0.401$  and  $-0.440$  at T1 and T2, respectively (all  $P < 0.001$ ), providing support for convergent validity. The Spearman correlations of the total PHQ-4-C scores with the total RU\_SATED-C scale scores were  $-0.221$  and  $-0.256$  at T1 and T2, respectively, (all  $P < 0.001$ ), providing support for divergent validity. The correlation matrix of the RU\_SATED-C scores on inter-item and item-total, and with the SQQ-C and the PHQ-4-C scores on subscales and global scale, are presented in Additional file 1: Table S2. Spearman correlations to establish convergent and divergent validity are shown in Fig. 1.

**Internal consistency and test-retest reliability**

Ordinal Cronbach’s alpha at T1 and T2 were 0.670 and 0.722, respectively. Ordinal McDonald’s omega at T1 and T2 were 0.676 and 0.725, respectively. Both metrics are suggestive of suboptimal levels of internal consistency (Table 4). ICC analyses showed that the RU\_SATED-C scale and items were significantly correlated across time

intervals (ICC = 0.354–0.683), suggestive of fair to good test-retest reliability, with the exception of item 5 which demonstrated poor test-retest reliability (Efficiency) (Table 4).

**Reliability for other measures**

Internal consistency and test-retest reliability of the SQQ-C and the PHQ-4-C were reported in Additional file 1: Table S3. Briefly, ordinal Cronbach’s alpha values ranged from 0.737 to 0.904, ordinal McDonald’s omega values ranged from 0.747 to 0.904; ICCs ranged 0.632 to 0.797 for the global scale, and its subscale. Regarding structural validity, some details of which were reported elsewhere [26, 55], the SQQ-C and the PHQ-4-C respectively exhibited stable two-factor solution and favorable fit indices.

**Discussion**

The aim of this study was to translate and adapt the RU\_SATED scale for use in Chinese and validate the RU\_SATED-C scale to provide preliminary reliability and validity when used for assessing sleep health in a Chinese population. The methodology used was similar to that used in the various languages’ validation studies of the

**Table 4** Internal consistency and test–retest reliability for the RU\_SATED-C at T1 and T2 (N = 898)

	Total/items	SEM	
		T1	T2
Ordinal Cronbach's Alpha			
T1	0.670 (0.637, 0.704)		
T2	0.722 (0.693, 0.750)		
Ordinal McDonald's Omega			
T1	0.676 (0.643, 0.708)		
T2	0.725 (0.698, 0.753)		
ICC (T1, T2)	0.683 (0.646, 0.716)	1.216	1.259
01. Do you go to bed and get out of bed at about the same time (within one hour) every day?	0.464 (0.411, 0.514)	0.403	0.406
02. Are you satisfied with your sleep?	0.621 (0.579, 0.660)	0.406	0.398
03. Do you stay awake all day without dozing?	0.561 (0.513, 0.604)	0.404	0.411
04. Is the middle of your sleep between 2:00 a.m. and 4:00 a.m.?	0.616 (0.574, 0.655)	0.506	0.476
05. Do you spend less than 30 min awake at night? This includes the time it takes to fall asleep plus awakenings during sleep.	0.354 (0.296, 0.410)	0.597	0.593
06. Do you sleep between 6 and 8 h per day?	0.531 (0.482, 0.576)	0.318	0.331

Point estimate and 95% confidence interval; Estimates on Cronbach's Alpha and McDonald's Omega assuming ordinal level; T1, Time 1; T2, Time 2; ICC, intraclass correlation coefficient; SEM, standard error of measurement; SEM = SD × sqrt(1-ICC); SD, standard deviation

RU\_SATED scale, in Portuguese (2018), English (2019), Spanish (2020), French (2021), and Japanese (2022) [9, 16–19]. Admittedly, this is the first study to assess the psychometric performance of the RU\_SATED in a Chinese population. This instrument demonstrated adequate measurement properties when used with Chinese health-care students.

Specifically, model fit indices produced by CFAs indicated that a single-factor structure fit the data well at two time points, similar to that of Portuguese, Spanish, and French studies [16–18]. Note, however, that some significant distinctions in translations exist. For example, the Portuguese version utilized a five-point Likert scale, and the Spanish version was adapted from the original five-item SATED scale. Our observed single-factor model differed from the two-factor model obtained in the English and Japanese studies [9, 19]. It is important to note that there are differences between the two-factor models obtained in English [Factor 1 (Sleep Quality & Quantity): Satisfaction, Efficiency, Duration; Factor 2 (Circadian Rhythm): Regularity, Alertness, Timing] and Japanese [Factor 1 (Quality & Quantity): Satisfaction, Alertness, Duration; Factor 2 (Circadian): Regularity, Efficiency, Timing]. In the French validation, the CFA showed an acceptable fit for both the one-factor and two-factor structures (in common with the English version); however, the fit was slightly better for the latter [18]. Our data were well approximated by a single-factor model across both testing occasions and also supported strict factorial invariance across time. The data supported the two-factor models found in the Japanese and English samples

and resulted in a better fit for the two-factor model previously found in the Japanese version and a slightly worse fit for the two-factor model found in the English version, respectively. One underlying reason is that both China and Japan belong to Oriental cultures, and they may have similar understandings of sleep health. However, there may be significant differences in the understanding of sleep habits between Oriental and Occidental cultures, such as their siesta habit and sleeping partners. Asians were found to perceive sleep problems more often than individuals of the Americas [56], perceived a weaker relation between sleep and physical health, and had a significantly shorter ideal amount of sleep [57]. More research is needed not only to replicate these studies, but to learn more about sleep health constructs.

Convergent and divergent validity were assessed with the SQQ-C and the PHQ-4-C, revealing satisfactory correlations, all in the expected directions. To establish convergent validity, ideally a multidimensional sleep scale would be used; however, since there is no scale that meets this requirement, a recently developed sleep quality questionnaire that is considered to partially overlap in terms of underlying constructs was adopted as a reference. The SQQ-C and the RU\_SATED-C scale total scores were moderately correlated, while the correlation between the PHQ-4-C and the RU\_SATED-C scale total scores was weak. For internal consistency, ordinal Cronbach's alpha between the two testing times was slightly better than that of the English (0.64) and French (0.57) versions [9, 18], and were considerably lower than that of the Japanese (0.758), Spanish (0.77), and Portuguese



(0.85) versions [16, 17, 19]. One potential reason behind such discrepancies is the small number of items (six) and three-point Likert-type response choices, which are known to decrease alpha. A second explanation for the lower internal consistency values might be related to the multifaceted nature of sleep health and cultural differences. The duration of sleep, sleeping locations, baby sleeping practices, ideology about napping, and more are all influenced by differences in cultures [56–59]. Cultural differences in sleep habits have a bearing on sleep health dimensions. Given some deficiencies of alpha, omega might be a practical alternative [60]. While no prior studies reported McDonald's omega, our results confirmed that omega values tended to outperform the alpha values.

The present translation, adaptation, and validation of the RU\_SATED scale for use in Chinese have notable strengths and limitations. First, our study provides initial evidence of the transcultural validation of the RU\_SATED scale with support in the form of acceptable psychometric performance of the RU\_SATED scale in Chinese, especially in terms of longitudinal measurement invariance and test–retest reliability. Second, the ordinal nature of item-level response choices was fully considered using McDonald's omega to evaluate internal consistency. Third, the RU\_SATED scale is a brief, simple, and versatile assessment tool—its translation and adaptation to Chinese represent an important step toward universal assessment of sleep health. In addition to strengths, several limitations need to be acknowledged. First, the lack of objective measures of sleep, specifically regarding sleep timing and efficiency, which are assessed with the RU\_SATED scale, may be considered potential limitations. However, objective measures of sleep can be impractical and expensive and thus infeasible for many large-scale studies. Although we did assess the convergent and divergent validity of the RU\_SATED scale by comparing scores to another sleep-related scale (e.g., SQQ) and a scale not assessing sleep (e.g., PHQ-4), future studies should examine the associations with other measures of sleep health (e.g., SHI). Second, the low internal consistency (Cronbach's alpha and McDonald's omega) and the short-interval test–retest are two limitations of the study, perhaps restricting its ability to open practical prospects. Adding item(s) about sleep health behaviors [8] and scoring changes need to be considered as well. Third, another potential problem with generalizability from this sample is the restricted age range (minimum 17 to maximum 31 years, median = 20, interquartile range = 1). Finally, only a single cohort of healthcare students was used in this validation study. Participants, given their training, had unique medical knowledge, which may have led decrease generalizability. Importantly, traditions, cultural values, and local conditions and environments can

influence sleep practices and attitudes. Therefore, future studies should further evaluate the measurement properties of the RU\_SATED-C scale in additional validation studies, such as validating in community residents or a nationally representative sample.

## Conclusion

We cross-culturally adapted and validated the RU\_SATED scale for use in Chinese-speaking samples. This represents an important step in continuing efforts to promote healthy sleep and confirms promising measurement properties including longitudinal measurement. The RU\_SATED-C scale appears to be an easy-to-use and valid instrument for the measurement of multidimensional sleep health in healthcare students. Use of the RU\_SATED-C scale may begin to raise awareness of sleep health and could pave the way for important efforts to promote healthy sleep.

## Abbreviations

CFA	confirmatory factor analysis
CFI	comparative fit index
DSS	Daytime Sleepiness Subscale
DWLS	mean and variance adjusted diagonally weighted least squares
GAD-2	Generalized Anxiety Disorder-2
ICC	intraclass correlation coefficient
LCFA	longitudinal confirmatory factor analysis
LMI	longitudinal measurement invariance
PHQ-4	Patient Health Questionnaire-4
PHQ-2	Patient Health Questionnaire-2
RMSEA	root means square error of approximation
RU_SATED	Regularity, Satisfaction, Alertness, Timing, Efficiency, Duration
SDS	Sleep Difficulty Subscale
SQQ	Sleep Quality Questionnaire
TLI	Tucker–Lewis index

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40359-023-01203-5>.

**Additional file 1: Table S1.** Means, standard deviations, skewness, kurtosis and missing of scores on items and the RU\_SATED-C scale at T1 and T2 (N = 911). **Table S2.** The correlation matrix of the RU\_SATED-C scores on inter-item and item-total, and with the SQQ-C and the PHQ-4-C scores on subscales and global scale (N = 911). **Table S3.** Internal consistency and test-retest reliability of the SQQ-C and the PHQ-4-C at T1 and T2 (N = 911).

## Acknowledgements

The authors appreciate Daniel J. Buysse (MD, Professor, Department of Psychiatry, University of Pittsburgh, USA) and Carolyn Weber (MBA, Licensing Associate, Innovation Institute, University of Pittsburgh, USA) for use permission of the RU\_SATED scale. The authors appreciate three reviewers and Editorial Board Members for their helpful suggestions. The authors thank Yi Luo (MSN, RN, Associate Professor, School of Nursing, Ningbo College of Health Sciences, China) for data collection, Tuo Liu (PhD Candidate, Institute of Psychology, Goethe University Frankfurt, Frankfurt am Main, Germany) for discussion on statistical analysis, and Chen Jiang (Master Candidate, School of Public Health, Hangzhou Normal University, China) for assistance in data visualization. Additionally, the authors thank all the study participants and the research assistants for their contributions.

### Author contributions

RM: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing—Original Draft, Writing—Review & Editing. LD, JMD, SM, and KS: Methodology, Validation, Writing—Review & Editing. MC, BY, JW, BG, and JL: Validation. All authors read and approved the final manuscript.

### Funding

This study was funded by the Medical Research Fund of Zhejiang Province, Grant No. 2023RC073 and the Research Initiation Fund of Hangzhou Normal University, Grant No. RWSK20201003.

### Availability of data and materials

Permission to use the RU\_SATED scale should be requested to Prof. Daniel J. Buysse. All rights related to the RU\_SATED-C scale are reserved by the University of Pittsburgh. All data generated and/or analyzed during this study are not publicly available due to restrictions imposed by the ethics committee. The dataset supporting the conclusions is available upon reasonable request to the corresponding author.

### Declarations

#### Ethics approval and consent to participate

The study has been reviewed and approved by the Institutional Review Board of Hangzhou Normal University Division of Health Sciences, China (Reference No. 20190076), thus ensuring that it adhered to the standards set by the Helsinki Declaration. All healthcare students freely consented to answer the questionnaires and provided their informed consent before their inclusion into the survey. The authors confirmed full respect and protection of individual privacy rights before, during and after the data collection and processing.

#### Consent for publication

Not applicable.

#### Competing interests

The study was conducted in the absence of any potential competing interests.

#### Author details

<sup>1</sup>School of Public Health, Hangzhou Normal University, Hangzhou 311121, Zhejiang, China. <sup>2</sup>Engineering Research Center of Mobile Health Management System, Ministry of Education, Hangzhou, Zhejiang, China. <sup>3</sup>Department of Behavioral and Policy Sciences, RAND Corporation, Santa Monica, CA, USA. <sup>4</sup>Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA. <sup>5</sup>The National Sleep Foundation, Washington, DC, USA. <sup>6</sup>Department of Psychology, University of Crete, Rethymno, Greece. <sup>7</sup>Department of Youth and Family, Utrecht University, Utrecht, The Netherlands. <sup>8</sup>Faculty of Psychology, Southwest University, Chongqing, China. <sup>9</sup>Student's Mental Health Center, Sichuan Technology and Business University, Chengdu, Sichuan, China. <sup>10</sup>School of International Journalism and Communication, Beijing Foreign Studies University, Beijing, China. <sup>11</sup>School of Basic Medical, Shanghai University of Traditional Chinese Medicine, Shanghai, China. <sup>12</sup>Department of Oncology, Taihe Hospital, Hubei University of Medicine, Shiyan, Hubei, China. <sup>13</sup>Department of Behavioral Sciences and Health Education, Rollins School of Public Health, Emory University, Atlanta, GA, USA. <sup>14</sup>Université de Paris, NeuroDiderot, INSERM, Paris, France.

Received: 6 October 2022 Accepted: 9 May 2023

Published online: 05 July 2023

### References

- Hale L, Troxel W, Buysse DJ. Sleep health: an opportunity for public health to address health equity. *Annu Rev Public Health*. 2020;41(1):81–99. <https://doi.org/10.1146/annurev-publhealth-040119-094412>.
- Ramar K, Malhotra RK, Carden KA, Martin JL, Abbasi-Feinberg F, Aurora RN, Kapur VK, Olson EJ, Rosen CL, Rowley JA, et al. Sleep is essential to health: an American Academy of Sleep Medicine position statement. *J Clin Sleep Med*. 2021;17(10):2115–9. <https://doi.org/10.5664/jcsm.9476>.
- Buysse DJ. Sleep health: Can we define it? Does it matter? *Sleep*. 2014;37(1):9–17. <https://doi.org/10.5665/sleep.3298>.
- Mead MP, Irish LA. Application of health behaviour theory to sleep health improvement. *J Sleep Res*. 2020;29(5):e12950. <https://doi.org/10.1111/jsr.12950>.
- Espie CA. The '5 principles' of good sleep health. *J Sleep Res*. 2022;31(3):e13502. <https://doi.org/10.1111/jsr.13502>.
- van de Langenberg SCN, Kocavska D, Luik AI. The multidimensionality of sleep in population-based samples: a narrative review. *J Sleep Res*. 2022;31(4):e13608. <https://doi.org/10.1111/jsr.13608>.
- Ferini-Strambi L, Auer R, Bjorvatn B, Castronovo V, Franco O, Gabutti L, Galbiati A, Hajak G, Khatami R, Kitajima T, et al. Insomnia disorder: clinical and research challenges for the 21st century. *European J Neurol*. 2021;28(7):2156–67. <https://doi.org/10.1111/ene.14784>.
- Meltzer LJ, Williamson AA, Mindell JA. Pediatric sleep health: it matters, and so does how we define it. *Sleep Med Rev*. 2021;57:101425. <https://doi.org/10.1016/j.smrv.2021.101425>.
- Ravvits SG, Dzierzewski JM, Perez E, Donovan EK, Dautovich ND. Sleep health as measured by RU SATED: a psychometric evaluation. *Behav Sleep Med*. 2019;19(1):48–56. <https://doi.org/10.1080/15402002.2019.1701474>.
- Knutson KL, Phelan J, Paskow MJ, Roach A, Whiton K, Langer G, Hillygus DS, Mokrzycki M, Broughton WA, Chokroverty S, et al. The national sleep foundation's sleep health index. *Sleep Health*. 2017;3(4):234–40. <https://doi.org/10.1016/j.sleh.2017.05.011>.
- Azad MC, Fraser K, Rumana N, Abdullah AF, Shahana N, Hanly PJ, Turin TC. Sleep disturbances among medical students: a global perspective. *J Clin Sleep Med*. 2015;11(01):69–74. <https://doi.org/10.5664/jcsm.4370>.
- Yu D, Ren Q, Dong B, Zhao D, Sun Y. The sleep quality of medical students in China: a meta-analysis. *Sleep Biol Rhythm*. 2017;15(4):299–310. <https://doi.org/10.1007/s41105-017-0118-6>.
- Jahrami H, Dewald-Kaufmann J, Faris Me A-I, AlAnsari AMS, Taha M, AlAnsari N. Prevalence of sleep problems among medical students: a systematic review and meta-analysis. *J Public Health*. 2020;28(5):605–22. <https://doi.org/10.1007/s10389-019-01064-6>.
- Rao W-W, Li W, Qi H, Hong L, Chen C, Li C-Y, Ng CH, Ungvari GS, Xiang Y-T. Sleep quality in medical students: a comprehensive meta-analysis of observational studies. *Sleep Breath*. 2020;24(3):1151–65. <https://doi.org/10.1007/s11325-020-02020-5>.
- Sun Y, Wang H, Jin T, Qiu F, Wang X. Prevalence of sleep problems among Chinese medical students: a systematic review and meta-analysis. *Front Psychiatr*. 2022;13:753419. <https://doi.org/10.3389/fpsy.2022.753419>.
- Brandolim Becker N, Martins RIS, Jesus SN, Chioldelli R, Stephen Rieber M. Sleep health assessment: a scale validation. *Psychiatr Res*. 2018;259:51–5. <https://doi.org/10.1016/j.psychres.2017.10.014>.
- Benítez I, Roure N, Pinilla L, Sapiña-Beltrán E, Buysse DJ, Barbé F, de Batlle J. Validation of the satisfaction, alertness, timing, efficiency and duration (SATED) questionnaire for sleep health measurement. *Ann Am Thorac Soc*. 2020;17(3):338–43. <https://doi.org/10.1513/AnnalsATS.201908-628OC>.
- Coelho J, Lopez R, Richaud A, Buysse DJ, Wallace ML, Philip P, Micoulaud-Franchi J-A. Toward a multi-lingual diagnostic tool for the worldwide problem of sleep health: the French RU-SATED validation. *J Psychiatr Res*. 2021;143:341–9. <https://doi.org/10.1016/j.jpsychires.2021.09.008>.
- Furihata R, Tateyama Y, Nakagami Y, Akahoshi T, Itani O, Kaneita Y, Buysse DJ. The validity and reliability of the Japanese version of RU-SATED. *Sleep Med*. 2022;91:109–14. <https://doi.org/10.1016/j.sleep.2022.02.014>.
- Acquadro C, Conway K, Giroulet C, Mear I. Linguistic validation manual for health outcome assessments. Lyon: MAPI Research Trust; 2012.
- Zhu Y, Jiang C, Yang Y, Dzierzewski JM, Spruyt K, Zhang B, Huang M, Ge H, Rong Y, Ola BA, et al. Depression and anxiety mediate the association between sleep quality and self-rated health in healthcare students. *Behav Sci*. 2023;13(2):82. <https://doi.org/10.3390/bs13020082>.
- Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. 5th ed. Oxford, UK: Oxford University Press; 2015.
- McKechnie D, Fisher M. Considerations when examining the psychometric properties of measurement instruments used in health. *Aust J Adv Nurs*. 2022;39(2):36–47. <https://doi.org/10.37464/2020.392.481>.
- Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures statistics and strategies for evaluation. *Control Clin Trials*.

- 1991;12(4 Supplement):S142–58. [https://doi.org/10.1016/S0197-2456\(05\)80019-4](https://doi.org/10.1016/S0197-2456(05)80019-4).
25. Kato T. Development of the sleep quality questionnaire in healthy adults. *J Health Psychol*. 2014;19(8):977–86. <https://doi.org/10.1177/1359105313482168>.
  26. Meng R, Kato T, Mastrotheodoros S, Dong L, Fong DYT, Wang F, Cao M, Liu X, Yao C, Cao J, et al. Adaptation and validation of the Chinese version of the sleep quality questionnaire. *Qual Life Res*. 2023;32(2):569–82. <https://doi.org/10.1007/s11136-022-03241-9>.
  27. Luo Y, Fei S, Gong B, Sun T, Meng R. Understanding the mediating role of anxiety and depression on the relationship between perceived stress and sleep quality among health care workers in the COVID-19 response. *Nat Sci Sleep*. 2021;13:1747–58. <https://doi.org/10.2147/NSS.S313258>.
  28. Meng R, Lau EYY, Spruyt K, Miller CB, Dong L. Assessing measurement properties of a simplified Chinese version of sleep condition indicator (SCI-SC) in community residents. *Behav Sci*. 2022;12(11):433. <https://doi.org/10.3390/bs12110433>.
  29. Spitzer RL, Williams JBW, Kroenke K. PHQPCSG: validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA*. 1999;282(18):1737–44. <https://doi.org/10.1001/jama.282.18.1737>.
  30. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics*. 2009;50(6):613–21. [https://doi.org/10.1016/S0033-3182\(09\)70864-3](https://doi.org/10.1016/S0033-3182(09)70864-3).
  31. Patient Health Questionnaire (PHQ) Screeners. <https://www.phqscreeners.com/>.
  32. Tierney N, Cook D. Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. *J Stat Softw*. 2023;105(1):1–31. <https://doi.org/10.18637/jss.v105.i07>.
  33. Korkmaz S, Goksuluk D, Zararsiz G. MVN: an R package for assessing multivariate normality. *R J*. 2014;6(2):151–62. <https://doi.org/10.32614/RJ-2014-031>.
  34. Rosseel Y. Lavaan: an R package for structural equation modeling. *J Stat Softw*. 2012;48(2):1–36. <https://doi.org/10.18637/jss.v048.i02>.
  35. semTools: Useful tools for structural equation modeling. R package version 0.5-5. <https://CRAN.R-project.org/package=semTools>.
  36. irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. <https://CRAN.R-project.org/package=irr>.
  37. ufs: A collection of utilities. R package version 0.5.2. <https://r-packages.gitlab.io/ufs>.
  38. Bennett DA. How can I deal with missing data in my study? *Aust N Z J Public Health*. 2001;25(5):464–9. <https://doi.org/10.1111/j.1467-842X.2001.tb00294.x>.
  39. Data Prep. [http://statwiki.gskination.com/index.php?title=Data\\_Prep](http://statwiki.gskination.com/index.php?title=Data_Prep).
  40. Kline RB. Principles and practice of structural equation modeling. New York, NY: Guilford publications; 2016.
  41. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–57. <https://doi.org/10.1007/s11136-018-1798-3>.
  42. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737–45. <https://doi.org/10.1016/j.jclinepi.2010.02.006>.
  43. Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods*. 2004;9(4):466–91. <https://doi.org/10.1037/1082-989X.9.4.466>.
  44. Li C-H. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods*. 2016;48(3):936–49. <https://doi.org/10.3758/s13428-015-0619-7>.
  45. Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001;66(4):507–14. <https://doi.org/10.1007/BF02296192>.
  46. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis: Pearson new international edition. 7th ed. London: Pearson Higher Education; 2014.
  47. Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, Williamson PR, Terwee CB. How to select outcome measurement instruments for outcomes included in a “core outcome set”—a practical guideline. *Trials*. 2016;17(1):449. <https://doi.org/10.1186/s13063-016-1555-2>.
  48. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Develop Rev*. 2016;41:71–90. <https://doi.org/10.1016/j.dr.2016.06.004>.
  49. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Model Multidiscip J*. 2007;14(3):464–504. <https://doi.org/10.1080/10705510701301834>.
  50. Nelemans SA, Meeus WHJ, Branje SJT, Van Leeuwen K, Colpin H, Verschueren K, Goossens L. Social anxiety scale for adolescents (SAS-A) short form: longitudinal measurement invariance in two community samples of youth. *Assessment*. 2019;26(2):235–48. <https://doi.org/10.1177/1073191116685808>.
  51. Zumbo BD, Gadermann AM, Zeisser C. Ordinal versions of coefficients alpha and theta for likert rating scales. *J Modern Appl Stat Methods*. 2007;6(1):21–9. <https://doi.org/10.22237/jmasm/1177992180>.
  52. Crutzen R, Peters G-JY. Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychol Rev*. 2017;11(3):242–7. <https://doi.org/10.1080/17437199.2015.1124240>.
  53. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–63. <https://doi.org/10.1016/j.jcm.2016.02.012>.
  54. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284–90. <https://doi.org/10.1037/1040-3590.6.4.284>.
  55. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, Schneider A, Brähler E. A 4-item measure of depression and anxiety: validation and standardization of the patient health questionnaire-4 (PHQ-4) in the general population. *J Affect Disord*. 2010;122(1):86–95. <https://doi.org/10.1016/j.jad.2009.06.019>.
  56. Sadeh A, Mindell J, Rivera L. “My child has a sleep problem”: a cross-cultural comparison of parental definitions. *Sleep Med*. 2011;12(5):478–82. <https://doi.org/10.1016/j.sleep.2010.10.008>.
  57. Cheung BY, Takemura K, Ou C, Gale A, Heine SJ. Considering cross-cultural differences in sleep duration between Japanese and Canadian university students. *PLoS ONE*. 2021;16(4):e0250671. <https://doi.org/10.1371/journal.pone.0250671>.
  58. Gentry R. Cultural considerations and sleep. In: Benuto LT, Gonzalez FR, Singer J, editors. *Handbook of cultural factors in behavioral health: a guide for the helping professional*. Cham: Springer International Publishing; 2020. p. 323–9. [https://doi.org/10.1007/978-3-030-32229-8\\_23](https://doi.org/10.1007/978-3-030-32229-8_23).
  59. Chen X, Wang R, Zee P, Lutsey PL, Javaheri S, Alcántara C, Jackson CL, Williams MA, Redline S. Racial/ethnic differences in sleep disturbances: the multi-ethnic study of atherosclerosis (MESA). *Sleep*. 2015;38(6):877–88. <https://doi.org/10.5665/sleep.4732>.
  60. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol*. 2014;105(3):399–412. <https://doi.org/10.1111/bjop.12046>.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

