BMJ Open

Sex-related differential item functioning of Jenkins Sleep Scale

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-074867
Article Type:	Original research
Date Submitted by the Author:	19-Apr-2023
Complete List of Authors:	Juhola, Juhani; TYKS Turku University Hospital, Physical and Rehabilitation medicine Arokoski, J. P. A.; Helsingin yliopistollinen Keskussairaala, Department of Physical and Rehabilitation Medicine; Helsingin yliopisto, Ervasti, Jenni; Finnish Institute of Occupational Health, Topeliuksenkatu 41 a A, FI-00250, Helsinki, Finland, Kivimäki, Mika; University of Helsinki Faculty of Medicine, ; University College London, Dep. of Epidemiology and Public Health Vahtera, Jussi; Turun Yliopisto, Dep Public Health; Finnish Institute of Occupational Health, Myllyntausta, Saana; University of Turku, Saltychev, Mikhail; Turku University Hospital and University ofTurku, Turku, Finland , Department of Physical and Rehabilitation Medicine
Keywords:	SLEEP MEDICINE, Psychometrics, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Sex-related differential item functioning of Jenkins Sleep Scale

Running head: DIF of Jenkins Sleep Scale

Juhani Juhola MD[1], Jari Arokoski MD, PhD[2], Jenni Ervasti[3], Mika Kivimäki PhD[3,4,5], Jussi Vahtera PhD, MD[6,7], Saana Myllyntausta[8], Mikhail Saltychev MD, PhD[1]

[1] Department of Physical and Rehabilitation Medicine, Turku University Hospital and University of Turku, Turku, Finland

[2] Department of Physical and Rehabilitation Medicine, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

[3] Finnish Institute of Occupational Health, Helsinki, Finland

[4] Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[5] Department of Epidemiology and Public Health, University College London Medical School, London, United Kingdom

[6] Department of Public Health, University of Turku, Turku, Finland

[7] Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland

[8] Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland

Address for correspondence:

Juhani Juhola, Department of Physical and Rehabilitation Medicine, Turku University Hospital, PO Box 52, FI-20521, Turku, Finland

E-mail: juhani.juhola@tyks.fi, Tel.: +358 40 738 9494, Fax: +358 2 313 3730

Keywords: sleep disorder; psychometrics; validity

ABSTRACT

Background

The Jenkins Sleep Scale (JSS) is one of the most used sleep questionnaires. No studies have focused on the psychometric properties of the JSS including differential item functioning (DIF) by applying item response theory. However, previous knowledge has suggested that sex-related differences in sleep and circadian rhythms may impact evaluation of sleep disorders.

Objective

To investigate the psychometric properties of the JSS focusing especially on the potential sex-related DIF by applying the item response theory.

Methods

Survey-based data from 77,967 participants of the Finnish Public Sector cohort study. Item response theory analysis was applied to assess the difficulty and discrimination properties of the JSS for the entire cohort and for both sexes separately.

Results

The mean age was 51.9 (13.1) years and 63,618 (82%) were women. The mean JSS total score was 6.4 (4.8) points. All four items demonstrated a slight shift towards higher severity of sleep difficulties. The discrimination estimates for all four items were moderate to high with moderate overall discrimination of 0.98 for the JSS composite score. For two items, the discrimination parameter was steeper in men, and for two other items the discrimination was steeper in women. The differences were significant, $p<0.001$ and the DIF was relatively uniform.

Conclusions

The JSS showed overall good psychometric properties such as difficulty and discrimination, among public sector employees. The JSS was able to discriminate people with different severity of sleep disturbances. However, when using the JSS, the respondents might slightly underestimate the severity of these disturbances. Also, the JSS may produce slightly different results when applied to men and women. Nevertheless, even though these sex-related differences were significant statistically, they are probably neglectable when applied to clinical situations.

Strengths and limitations

- To our knowledge there is no other studies focused on the psychometric properties of the JSS assessed by applying item response theory or Rasch analysis.

- The generalizability of the results might be weakened by the sex disbalance of the studied cohort (women were predominated) as fewer men work in the public sector in Finland. However, with almost 15 000 men in our data, it is unlikely that this is a source of a major bias.
- The mean age of study participants was 52 years and, therefore, the results describe principally people in the last third of their working life span.
- The response rate was 70% and there was no analysis of whether the non-respondents' demographic characteristics might affect the results.

Contributors All the authors (JJ, SM, JPAA, JE, MS, MK and JV) substantially contributed to the conception and design of the work, drafting the work and revised it critically for important intellectual content, interpreted the data, and finally approved the version published. All the authors achieved an agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. JJ was responsible for preparing the first draft. MS was responsible for the main data analysis. JV and MK were responsible for the acquisition of data. JV was a guarantor.

Funding This study was supported by funding granted by the Academy of Finland (grant 633666 to MK); NordForsk (to MK and JV); the UK MRC (grant K013351 to MK) and The Finnish Environment Fund (grants 190172 and 118060 to SM).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval The ethics committee of the Hospital District of Helsinki and Uusimaa has approved the study.

Data availability statement Data are available upon reasonable request. Individual-level survey data cannot be made publicly available, but information on the data and analyses is available upon request to the corresponding author.

INTRODUCTION

There are numerous scales to assess the severity of sleep difficulties [1]. Patient-reported outcome measures (PROMs) are easy to use and cost-efficient means to detect and grade sleep disturbances. The Jenkins Sleep Scale (JSS) has been developed as a brief and standardized questionnaire for sleep disturbances in 1998[2]. It is one of the most used questionnaires in epidemiological studies[1-4]. The JSS has been translated into several languages[5-10] and its psychometric properties have been found to be both valid and reliable across different patient groups, such as patients with rheumatoid [7] and psoriatic arthritis [6] and ankylosing spondylitis [5], fibromyalgia [10 11], chest pain [12], and post cardiac surgery patients [2]. Only a few studies have evaluated the psychometric properties of the JSS in large non-clinical populations [2 3 8 13 14].

Previous studies have found the JSS to be internally consistent amongst patients with fibromyalgia[10 11], rheumatoid arthritis [7], ankylosing spondylitis [5] and psoriatic arthritis [6], as indicated by a Cronbach's alpha ranging from 0.7 to 0.9. Several studies have assessed the internal consistency of the JSS in a general population reporting respectively good to excellent Cronbach's alpha between 0.8 and 0.9 [2 3 8 9 13 14]. A few previous studies have assessed the factor structure of the JSS and observed the JSS to be a unidimensional scale [3 8 9]. The construct structure of the JSS analysis has also been assessed by a confirmatory factor analysis, showing strong correlations between all four items and a common factor [3]. So far, no studies have focused on the psychometric properties of the JSS assessed by applying item response theory or Rasch analysis. Item response a theory investigates the relationship between the performance on a test item and the average (in a particular population of interest) level of the ability that item was designed to measure. It does not assume that each item is equally difficult, where difficulty is understood as the level of measurable ability needed to get a particular response to an item. This differentiates item response theory from other methods, which assume equality of response difficulties when several items are measured on an ordinal scale. The item response theory suggests that these differences between item difficulties may clinically be relevant and they should be taken into account when interpreting the results obtained from a test with multiple items. Additionally, item response theory suggests that individual items, as well as an entire test, may perform differently in at the different levels of measurable ability.

Sex-related differences of sleep and circadian rhythm are well known [15 16]. It has been suggested that such differences may be age-related and occur as early as the middle years of life [15]. Even though women may have better sleep quality than men in relation to sleep length, sleep-onset latency and sleep efficiency [17], women have 1.5 times higher risk to develop insomnia than men, and this predisposition has been found to be consistent and progressive with ageing [18 19]. Shorter circadian cycle lengths as well as a larger amplitude of circadian variation in women may lead to more frequent nighttime impairment in women[16]. Sex differences in the sleep disorders underscore the need to account for sex in sleep medicine and sleep research [16]. Additionally, the diagnostics of diseases related to sleep disorders may differ between men

and women. For example, narcolepsy or sleep apnea may be diagnosed later (or even remained undiagnosed) in women, at least partially due to variation in presenting symptoms [16 20]. Restless legs syndrome is also more common among women [16 17 21]. It has also been suggested that the decreased need for sleep is associated with ageing (shorter sleep duration and nighttime awakenings) may be more common among men than women [22]. Sex differences in the incidence of insomnia are the result of a complex combination of biopsychosocial factors changing across the life span [16-18 20-26]. These differences may be related to sex hormones or specific sex-dependent patterns of physiological periods like puberty, menstruation, pregnancy, and menopause or other causes [16-18 20-26]. Most of previous studies on the topic have focused on sex- and age-related differences in prevalence and incidence of insomnia, while milder sleep disturbances have less been studied.

Previous studies have suggested that both sex differences in sleep and circadian rhythms may impact evaluation of sleep disorders [16]. Sleep scales, including the JSS, may possibly perform differently across sexes [24]. For example, it has been reported that women may perceive nighttime awakenings as more difficult than men [22]. The potential sex-related differential item functioning (DIF) of the JSS has not been studied before. The aim of this study was to investigate the psychometric properties of the JSS focusing especially on the potential sex-related DIF by applying the item response theory.

METHODS

The Finnish Public Sector (FPS) study cohort consists of employees in 11 town municipalities and five social- and health organizations.. The data were sourced from the survey in 2016 – 2017 (average response rate 70%). All the participants 77, 967, who have responded to at least one item of the JSS have been included. Individual-level survey data cannot be made publicly available, but information on the data and analyses are available upon request to the corresponding author. The ethics committee of the Hospital District of Helsinki and Uusimaa has approved the study (HUS 1210/2016).

All the data were obtained from the survey responses. Age was defined in full years at the time of survey response. Body mass index (BMI) was defined as weight divided by height to the power of two. The level of physical activity was calculated from the survey responses and converted into metabolic equivalent of task per hour per week (MET-h/week). Alcohol consumption was obtained from the survey and converted into gram/week. The respondents were asked about their usual amount of sleep hours per 24 hours with the following nine response alternatives: ≤6 hours, 6.5 hours, 7 hours, 7.5 hours, 8 hours, 8.5 hours, 9 hours, 9.5 hours, and ≥10 hours. The responses were then dichotomized as ≤7 hours vs. >7 hours of sleep.

The JSS is a four-item questionnaire used to grade the frequency of common sleep problems during the previous month [2]: trouble falling asleep, waking up but no trouble falling asleep again, waking up and trouble falling asleep again, waking up feeling tired (i.e., waking up after the usual amount of sleep feeling tired and worn out). Each item was rated on a Likert-like scale from zero to five, where zero is "never", one is "1-3 days", two is "about 1 night/week", three is "2-4 nights/week", four is "5-6 nights/week" and five is "almost every night". The total score is a simple sum of all four items' scores and ranges from zero ("no sleep problems") to 20 ("most sleep problems"). A score =<11 was considered as "little or no sleep disturbances" and >11 was considered as "high frequency of sleep disturbances" [27].

Statistical analysis

The results are reported as absolute numbers and percentages or as means and standard deviations (SD). The results were accompanied by 95% confidence intervals (95% CI) or two-tailed p-values, when appropriate. Using item response theory (IRT), the average level of reported sleep problems in the studied population was estimated based on the principle of maximum likelihood. Then, the level of sleep problems reported by each participant was compared to the average level observed in the entire sample. After fitting the model, both parameters – 'difficulty' and 'discrimination' – were calculated for each of the four items of the JSS by using the graded response model (GRM). Difficulty is the level of reported sleep problems needed to choose a particular response. In turn, discrimination is the steepness of the regression curve, with the severity of sleep problems placed on the X-axis and the expected score of the JSS on the Y-axis. Ideally, the steepest interval should correspond to the patients who obtained an average

score of two or three. If such is the case, then a test (or an item) is especially sensitive in distinguishing people with a level of sleep problems below average from those with levels above average. In this study, discrimination of 0.01 to 0.34 was considered 'none' (a totally level regression curve) or ‘very low’; 0.35 to 0.64 was considered 'low'; 0.65 to 1.34 was considered 'moderate'; 1.35 to 1.69 was considered 'high'; and a discrimination >1.7 was considered 'perfect' (a regression curve approaching a vertical line)[28]. An item information curve helps to comprehend this graphically, appointing the steepest interval of the curve to the level of disability that is associated with the most information that can be obtained from the item. Item information is calculated as an invert standard error. Results were reported along with their 95% confidence intervals (95%CI). The item characteristic curves of all four items are available from the corresponding author on request.

Differential item functioning (DIF) is a statistical characteristic of a scale item (here counted for each of four items included in JSS) that describes if the item is measuring an ability (here severity of sleep problems) differently for separate subgroups (here sexes) within the sample. To assess a DIF, the probit logistic regression was used to test whether an item exhibits either uniform or nonuniform DIF between sex groups, that is, whether an item favors one group over the other for all values of severity of sleep problems or for only some values of that [29 30]. A uniform DIF occurs when the difference between groups remains the same across the entire scale. In turn, a nonuniform DIF is observed when the direction of difference between groups varies at different levels of sleep problems (e.g., if men perform better than women up to a midpoint and worse than women after that). A two-tailed p-value ≤0.05 indicated a significant difference between sexes. When a significant DIF was observed, the results of DIF analysis were also presented and evaluated graphically as item information function curves. An item information function describes the precision, which an item or the entire test achieves for different levels of sleep difficulties. To put it in a simpler way, an item information function is an inverse variance.

The analyses were performed using Stata/IC Statistical Software: Release 17. College Station (StataCorp LP, TX, USA).

RESULTS

Of 81,136 participants, 77,967 have responded to at least one item of the JSS. 14,349 (18%) were men and 63,618 (82%) were women (Table 1). Their mean age was 51.9 (SD 13.1) years, body mass index 26.2 (SD 4.7) kg/m^2, physical activity 29.6 (25.3) METs/week, and alcohol consumption was 50.1 (SD 91.3) g/week (equivalent to around four units of alcohol per week). Of the respondents, 56,014 (72%) were sleeping seven or less hours per night. The mean JSS total score was 6.4 (SD 4.8) points. Of the respondents, 12,629 (16%) had JSS total score more than 11.

Difficulty parameter of JSS

Table 2 shows the estimates of difficulty parameter for all four items of the JSS. All four items demonstrated a slight shift towards higher severity of sleep difficulties – the estimates close to zero could be seen at the lowest end (instead of the middle point) of the scale. In other words, the respondents tended to underestimate their sleep difficulties. For example, for the item ”trouble falling asleep”, the respondents with slightly worse than average sleep difficulties still tended to mark the minimal possible score of one point. This shift towards underestimation was, however, mild. The same mild shift toward underestimation of the sleep problems was seen for both sexes (Table 3 and Figure 1).

Discrimination parameter of JSS

The discrimination estimates for the item “Waking up and trouble falling asleep again” were high for both sexes: 1.92 for men and 2.04 for women (Table 3). For the other three items, the estimates were moderate ranging in both sexes from 0.71 to 1.16. The overall discrimination of the composite JSS score was moderate 0.98 (95% CI 0.97 to 0.99).

Differential item functioning (DIF) of JSS

When considering both discrimination and difficulty parameters, there were significant differences between sexes, $p<0.001$. Figure 1 shows a test characteristic curve for the entire cohort. Figure 2 presents the item information functions of each item grouped by sex. For every JSS item and for both sexes, the most information could be observed at the slightly elevated levels of sleep disturbances. As shown in Figure 2, the discrimination parameter was steeper in men for the JSS items ‘Trouble falling asleep’ and ‘Waking up feeling tired’. Respectively, the discrimination was steeper in women for the items ‘Waking up but no trouble falling sleep again’ and ’Waking up and trouble falling asleep again’. The shapes of curves were close to uniform for all the items.

DISCUSSION

In this survey-based cross-sectional cohort study among 77,967 employees of the public sector, there were little differences in the psychometric properties of JSS between sexes. All four items demonstrated a slight shift towards higher severity of sleep difficulties; the respondents tended (but only mildly) to underestimate their sleep difficulties. This shift was seen for both sexes. The discrimination estimates ranged from moderate to high, which means that the JSS is a sensitive scale for distinguishing people with different levels of sleep difficulties. A uniform DIF (slight but statistically significant) was present for all four items; the JSS was more sensitive among men for items "Trouble falling asleep" and "Waking up feeling tired" and among women for items "Waking up but no trouble falling asleep again" and "Waking up and trouble falling asleep again". These differences may be related to different sleep disorders and to the differences of the incidence of these disorders between men and women. Women have more sex hormone related sleep disorders[16-18 20-26] and also restless leg syndrome [17], while men have more obstructive sleep apnea and sleep disordered breathing, which are known to cause trouble falling asleep but also increasing daytime tiredness [16 20]. These sex-related differences are that undeniable that although the behavioral treatments for insomnia have equal effects regardless of gender, certain pharmacologic treatments require different dosing based on sex [16].

The generalizability of the results might be weakened by the sex disbalance of the studied cohort (women were predominated) as fewer men work in the public sector in Finland. However, with almost 15 000 men in our data, it is unlikely that this is a source of a major bias. Also, the mean age of study participants was 52 years and, therefore, the results describe principally people in the last third of their working life span. While it has been widely used for over two decades, the Finnish translation of JSS had never undergone a full linguistic validation process, which might affect its equivalency with an English version. The response rate was 70% and there was no analysis of whether the non-respondents' demographic characteristics might affect the results.

The direct comparison between the present results and previous research is limited by the fact that no earlier studies have focused on the psychometric properties of the JSS assessed by applying item response theory or Rasch analysis. This might leave the following clinically relevant questions unanswered: does a Likert-like scale used by the JSS behave similarly for all four items, does the JSS (as an entire test and its individual items) perform differently across the whole severity spectrum of sleep disturbances and does the JSS perform equally well in diverse subgroups and situations? As such, this also was the first study exploring the DIF of the JSS. However, the results of this study reflect previously observed differences in the amount and severity of sleep difficulties among men and women [16 17 20-26]. The results are also in line with previously reported differences in the way men and women grade their sleep difficulties when responding to questionnaires [16 24]. Previous studies have suggested that both sex and gender differences

in sleep and circadian rhythms may impact evaluation of sleep disorders and some sleep scales, including the JSS, may behave differently among men and women [16]. For example, the Pittsburgh Sleep Quality Index has showed similar sex-related inconsistency [31]. The DIF has been detected for the Karolinska Sleep Questionnaire [32]. Also, the Patient-Reported Outcomes Measurement Information System (PROMIS), which is a very popular standard general patient-reported outcome measure, has demonstrated an age-related DIF for a sleep- item [33].

The sex-related DIF of patient-reported outcome measures is a common finding. For example, such a DIF has been found for scales measuring quality of life, depression, disability caused by pain and general disability [34-40]

Simply explained, the significance of the results from a clinical point of view is that the JSS works almost equally well across both sexes. The DIF observed here was small and uniform, hardly affecting the interpretation or comparison of the scores obtained from a few individual people. On the other hand, this DIF may be of significant importance when the JSS is used to collect data from large populations, especially when comparing populations with dissimilar sex distributions. If there is such a situation, then that comparison should separately be performed by sex groups. This can be particularly true when, in addition to a composite score, research question concerns scores obtained from the JSS individual items.

Further research may reveal the potential DIF of the JSS among people of different age groups working in other fields than public sector, assuming that diverse physical and psychological work demands might affect the results obtained by the JSS. In addition, populations with different comorbidities (e.g., sleep apnea and disordered breathing or cardiovascular and metabolic disorders) may show results, which are different to the present ones.

Conclusions

The JSS showed overall good psychometric abilities, such as difficulty and discrimination, among public sector employees. The JSS was able to discriminate people with different severity of sleep disturbances. However, when using the JSS, the respondents might slightly underestimate the severity of these disturbances. Also, the JSS may produce slightly different results when applied to men or women. Nevertheless, even though these sex-related differences were statistically significant, they are probably neglectable when applied to clinical situations.

TABLES AND FIGURES

Table 1. Descriptive characteristics of the study cohort

Variable	Men		Women		Entire cohort	
	N	%	N	%	N	%
Total					77,967	100
Sex						
Men	100	0	0	100	14,349	18
Women	0	100	100	0	63,618	82
Sleep						
≤7 hours/night	10,779	75	45,235	71	56,014	72
>7 hours/night	3,570	25	18,383	29	21,953	28
	Mean	SD	Mean	SD	Mean	SD
Age, years	52.5	13.3	51.7	13.1	51.9	13.1
Physical activity, MET-hour/week	33.3	29.9	28.8	24.1	29.6	25.3
Body mass index, kg/m^2	26.9	4.0	26.1	4.8	26.2	4.7
Alcohol consumption, grams	99.4	156.0	38.9	63.7	50.1	91.3
Jenkin's Sleep Scale, points	5.8	4.6	6.6	4.8	6.4	4.8

Table 2. Difficulty coefficients of the JSS items in both sexes together (n=77,967)

Items and scores	Difficulty	95% CI	
Trouble falling asleep			
1 vs. 0	0.18	0.16	0.19
2 vs. 1	1.19	1.18	1.21
3 vs. 2	0.85	0.83	0.86
4 vs. 3	2.66	2.64	2.69
5 vs. 4	1.31	1.28	1.33
Waking up but no trouble falling asleep again			
1 vs. 0	-0.79	-0.81	-0.78
2 vs. 1	0.22	0.20	0.24
3 vs. 2	-0.13	-0.14	-0.11
4 vs. 3	1.69	1.67	1.72
5 vs. 4	0.33	0.31	0.36
Waking up and trouble falling asleep again			
1 vs. 0	-0.23	-0.25	-0.22
2 vs. 1	0.78	0.77	0.80
3 vs. 2	0.44	0.42	0.45
4 vs. 3	2.25	2.23	2.28
5 vs. 4	0.90	0.87	0.92
Waking up feeling tired			
1 vs. 0	-0.54	-0.55	-0.52
2 vs. 1	0.48	0.46	0.50
3 vs. 2	0.13	0.12	0.15
4 vs. 3	1.95	1.93	1.98
5 vs. 4	0.59	0.57	0.62

Table 3. Difficulty and discrimination coefficients of the JSS items by sex

Items and scores	Men			Women		
	Coefficient	95% CI		Coefficient	95% CI	
Trouble falling asleep						
Discrimination	0.95	0.91	0.99	0.85	0.82	0.87
Difficulty						
1 vs. 0	0.41	0.36	0.47	0.42	0.39	0.45
2 vs. 1	1.22	1.15	1.29	1.29	1.25	1.34
3 vs. 2	1.26	1.17	1.34	1.11	1.07	1.16
4 vs. 3	3.05	2.87	3.23	3.24	3.14	3.34
5 vs. 4	1.03	0.84	1.23	1.04	0.95	1.13
Waking up but no trouble falling asleep again						
Discrimination	0.99	0.94	1.04	1.17	1.13	1.21
Difficulty						
1 vs. 0	-0.69	-0.75	-0.64	-0.76	-0.80	-0.72
2 vs. 1	0.60	0.53	0.67	0.60	0.56	0.63
3 vs. 2	0.05	-0.02	0.11	0.00	-0.03	0.03
4 vs. 3	1.84	1.74	1.94	1.76	1.70	1.81
5 vs. 4	0.21	0.11	0.31	0.35	0.31	0.40
Waking up and trouble falling asleep again						
Discrimination	1.92	1.81	2.04	2.04	1.96	2.11
Difficulty						
1 vs. 0	-0.10	-0.14	-0.07	0.04	0.02	0.07
2 vs. 1	0.58	0.54	0.61	0.67	0.64	0.70
3 vs. 2	0.62	0.58	0.66	0.66	0.63	0.68
4 vs. 3	1.65	1.59	1.72	1.70	1.65	1.74
5 vs. 4	1.34	1.27	1.40	1.38	1.34	1.42
Waking up feeling tired						
Discrimination	0.84	0.80	0.84	0.72	0.70	0.74
Difficulty						
1 vs. 0	-0.50	-0.56	-0.50	-0.53	-0.57	-0.49
2 vs. 1	0.61	0.55	0.61	0.73	0.69	0.77
3 vs. 2	0.43	0.36	0.43	0.20	0.16	0.24
4 vs. 3	2.26	2.14	2.26	2.40	2.33	2.48
5 vs. 4	1.06	0.93	1.06	0.94	0.88	1.01

Figure 1. Test characteristic curve in both sexes together

Figure 2. Item information functions of JSS items grouped by sex

References

1. Lallukka T, Dregan A, Armstrong D. Comparison of a sleep item from the General Health Questionnaire-12 with the Jenkins Sleep Questionnaire as measures of sleep disturbance. *J Epidemiol* 2011;21(6):474-80. doi: 10.2188/jea.je20110023 [published Online First: 2011/10/12]
2. Jenkins CD, Stanton BA, Niemcryk SJ, et al. A scale for the estimation of sleep problems in clinical research. *J Clin Epidemiol* 1988;41(4):313-21. doi: 10.1016/0895-4356(88)90138-2 [published Online First: 1988/01/01]
3. Tibubos AN, Zenger M, Schmalbach B, et al. Measurement invariance, validation and normative data of the Jenkins Sleep Scale-4 (JSS-4) in the German general population across the life span. *J Psychosom Res* 2020;130:109933. doi: 10.1016/j.jpsychores.2020.109933 [published Online First: 2020/01/18]
4. Vahtera J, Pentti J, Helenius H, et al. Sleep disturbances as a predictor of long-term increase in sickness absence among employees after family death or illness. *Sleep* 2006;29(5):673-82. doi: 10.1093/sleep/29.5.673 [published Online First: 2006/06/16]
5. Duruoz MT, Ulutatar F, Ozturk EC, et al. Assessment of the validity and reliability of the Jenkins Sleep Scale in ankylosing spondylitis. *Int J Rheum Dis* 2019;22(2):275-79. doi: 10.1111/1756-185x.13447 [published Online First: 2018/12/20]
6. Duruöz MT, Erdem D, Gencer K, et al. Validity and reliability of the Turkish version of the Jenkins Sleep Scale in psoriatic arthritis. *Rheumatology International* 2018;38:261-65.
7. Duruöz MT, Ünal Ç, Ulutatar F, et al. The Validity and Reliability of Turkish Version of the Jenkins Sleep Evaluation Scale in Rheumatoid Arthritis. *Arch Rheumatol* 2018;33(2):160-67. doi: 10.5606/ArchRheumatol.2018.6376 [published Online First: 2018/09/13]
8. Reis C, Mestre C, Tecedeiro MMV, et al. Translation, cross-cultural adaptation and psychometric properties of the Jenkins Sleep Scale in a sample of Portuguese shift workers. *Laboratório de Psicologia* 2014;12:89-98.
9. Salman M, Khan AH, Sulaiman SAS, et al. Psychometric validation of the Urdu version of the Jenkins sleep evaluation questionnaire. *Journal of Pharmacy Practice and Community Medicine* 2018;4(1)
10. Unal-Ulutatar C, Ozsoy-Unubol T. Psychometric properties of Turkish version of Jenkins sleep scale in fibromyalgia syndrome. *Advances in Rheumatology* 2020;60
11. Crawford B, Piault E, Lai C, et al. Assessing sleep in fibromyalgia: investigation of an alternative scoring method for the Jenkins Sleep Scale based on data from randomised controlled studies. *Clinical and Experimental Rheumatology-Incl Supplements* 2010;28(6):S100.
12. Jerlock M, Gaston-Johansson F, Kjellgren KI, et al. Coping strategies, stress, physical activity and sleep in patients with unexplained chest pain. *BMC nursing* 2006;5(1):1-10.
13. Nasermoaddeli A, Sekine M, Kumari M, et al. Association of sleep quality and free time leisure activities in Japanese and British civil servants. *Journal of occupational health* 2005;47(5):384-90.
14. Ornat L, Martínez-Dearth R, Chedraui P, et al. Assessment of subjective sleep disturbance and related factors during female mid-life with the Jenkins Sleep Scale. *Maturitas* 2014;77(4):344-50.
15. Carrier J, Semba K, Deurveilher S, et al. Sex differences in age-related changes in the sleep-wake cycle. *Frontiers in neuroendocrinology* 2017;47:66-85.
16. Marsella JL, Sharkey KM. Sex Differences in Sleep Disorders. In: Attarian H, Viola-Saltzman M, eds. Sleep Disorders in Women: A Guide to Practical Management. Cham: Springer International Publishing 2020:65-81.
17. Krishnan V, Collop NA. Gender differences in sleep disorders. *Curr Opin Pulm Med* 2006;12(6):383-9. doi: 10.1097/01.mcp.0000245705.69440.6a [published Online First: 2006/10/21]
18. Suh S, Cho N, Zhang J. Sex Differences in Insomnia: from Epidemiology and Etiology to Intervention. *Curr Psychiatry Rep* 2018;20(9):69. doi: 10.1007/s11920-018-0940-9 [published Online First: 2018/08/11]
19. Zhang B, Wing Y-K. Sex Differences in Insomnia: A Meta-Analysis. *Sleep* 2006;29(1):85-93. doi: 10.1093/sleep/29.1.85

20. Bublitz M, Adra N, Hijazi L, et al. A Narrative Review of Sex and Gender Differences in Sleep Disordered Breathing: Gaps and Opportunities. *Life* 2022;12(12):2003.
21. Meers J, Stout-Aguilar J, Nowakowski S. Chapter 3 - Sex differences in sleep health. In: Grandner MA, ed. Sleep and Health: Academic Press 2019:21-29.
22. Jonasdottir SS, Minor K, Lehmann S. Gender differences in nighttime sleep patterns and variability across the adult lifespan: a global-scale wearables study. *Sleep* 2020;44(2) doi: 10.1093/sleep/zsaa169
23. Madrid-Valero JJ, Kirkpatrick RM, González-Javier F, et al. Sex differences in sleep quality and psychological distress: Insights from a middle-aged twin sample from Spain. *Journal of Sleep Research*;n/a(n/a):e13714. doi: https://doi.org/10.1111/jsr.13714
24. Mallampalli MP, Carter CL. Exploring sex and gender differences in sleep health: a Society for Women's Health Research Report. *J Womens Health (Larchmt)* 2014;23(7):553-62. doi: 10.1089/jwh.2014.4816 [published Online First: 2014/06/24]
25. Mong JA, Baker FC, Mahoney MM, et al. Sleep, rhythms, and the endocrine brain: influence of sex and gonadal hormones. *J Neurosci* 2011;31(45):16107-16. doi: 10.1523/jneurosci.4175-11.2011 [published Online First: 2011/11/11]
26. Mong JA, Cusmano DM. Sex differences in sleep: impact of biological sex and sex steroids. *Philos Trans R Soc Lond B Biol Sci* 2016;371(1688):20150110. doi: 10.1098/rstb.2015.0110 [published Online First: 2016/02/03]
27. Monterrosa-Castro Á, Portela-Buelvas K, Salguedo-Madrid M, et al. Instruments to study sleep disorders in climacteric women. *Sleep Science* 2016;9(3):169-78.
28. Baker FB. The basics of item response theory. Madison, WI, USA: ERIC Clearinghouse on Assessment and Evaluation 2001.
29. De Boeck P, Wilson M. Explanatory item response models: A generalized linear and nonlinear approach: Springer 2004.
30. Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. *Journal of Educational measurement* 1990;27(4):361-70.
31. Mollayeva T, Thurairajah P, Burton K, et al. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. *Sleep Med Rev* 2016;25:52-73. doi: 10.1016/j.smrv.2015.01.009 [published Online First: 2015/07/15]
32. Santos KD, Fischer FM, de Andrade DF, et al. Shift work and sex are variants across Karolinska Sleep Questionnaire and the application of linking analysis as a step forward using Item Response Theory. *Chronobiol Int* 2019;36(1):42-56. doi: 10.1080/07420528.2018.1515773 [published Online First: 2018/11/08]
33. van Kooten J, Terwee CB, Luijten MAJ, et al. Psychometric properties of the Patient-Reported Outcomes Measurement Information System (PROMIS) Sleep Disturbance and Sleep-Related Impairment item banks in adolescents. *J Sleep Res* 2021;30(2):e13029. doi: 10.1111/jsr.13029 [published Online First: 2020/03/18]
34. Knott RJ, Lorgelly PK, Black N, et al. Differential item functioning in quality of life measurement: An analysis using anchoring vignettes. *Soc Sci Med* 2017;190:247-55. doi: 10.1016/j.socscimed.2017.08.033 [published Online First: 2017/09/08]
35. Saltychev M, Widbom-Kolhanen SS, Pernaa KI. Sex-related differential item functioning of neck disability index. *Disabil Rehabil* 2023:1-7. doi: 10.1080/09638288.2023.2180545 [published Online First: 2023/02/22]
36. Katajapuu N, Laimi K, Heinonen A, et al. Floor and ceiling effects of the World Health Organization Disability Assessment Schedule 2.0 among patients with chronic musculoskeletal pain. *Int J Rehabil Res* 2019;42(2):190-92. doi: 10.1097/MRR.0000000000000339 [published Online First: 2019/02/15]
37. Bourion-Bedes S, Schwan R, Laprevote V, et al. Differential item functioning (DIF) of SF-12 and Q-LES-Q-SF items among french substance users. *Health Qual Life Outcomes* 2015;13:172. doi: 10.1186/s12955-015-0365-7 [published Online First: 2015/10/27]

38. Cavanagh A, Wilson CJ, Caputi P, et al. Symptom endorsement in men versus women with a diagnosis of depression: A differential item functioning approach. *Int J Soc Psychiatry* 2016;62(6):549-59. doi: 10.1177/0020764016653980 [published Online First: 2016/06/24]
39. Verdam MGE, Oort FJ, Sprangers MAG. Item bias detection in the Hospital Anxiety and Depression Scale using structural equation modeling: comparison with other item bias detection methods. *Qual Life Res* 2017;26(6):1439-50. doi: 10.1007/s11136-016-1469-1 [published Online First: 2016/12/13]
40. Brzoska P. Assessment of quality of life in individuals with chronic headache. Psychometric properties of the WHOQOL-BREF. *BMC Neurol* 2020;20(1):267. doi: 10.1186/s12883-020-01845-7 [published Online First: 2020/07/06]

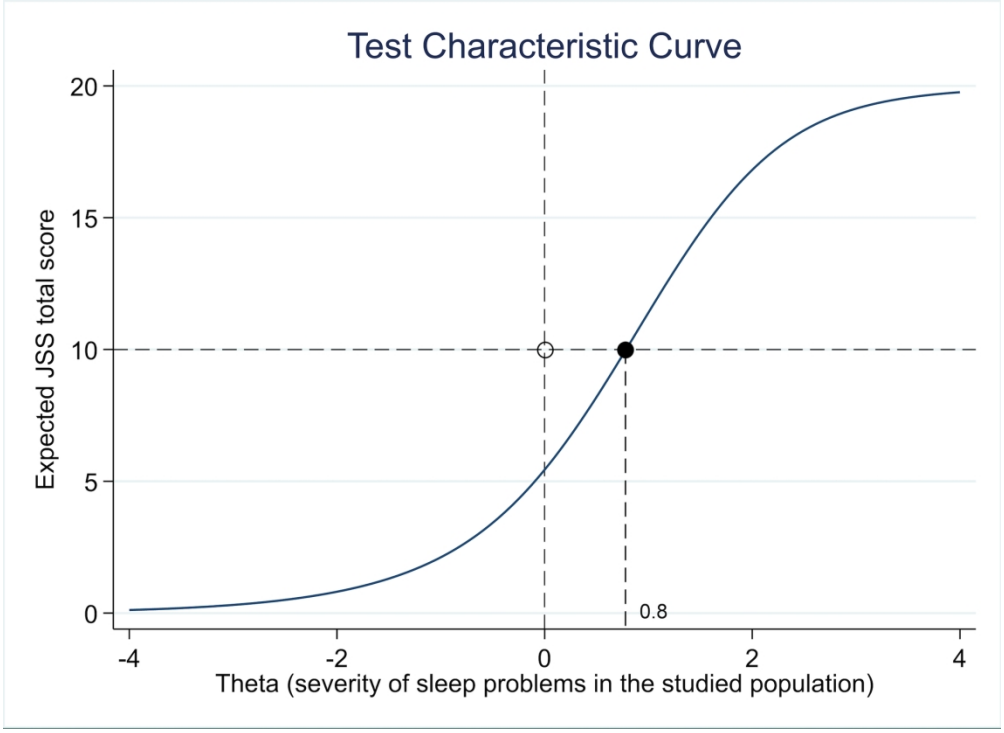

Figure 1. Test characteristic curve in both sexes together

87x63mm (600 x 600 DPI)

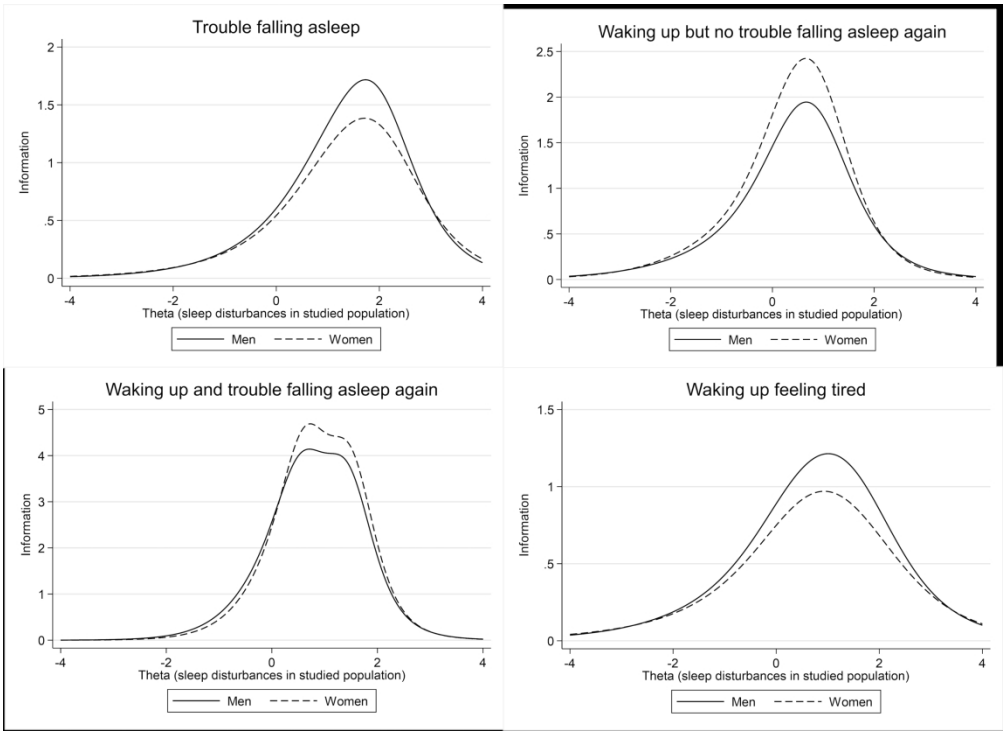

Figure 2. Item information functions of JSS items grouped by sex

147x107mm (600 x 600 DPI)

Reporting checklist for qualitative study.

Based on the SRQR guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the SRQRreporting guidelines, and cite them as:

O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med. 2014;89(9):1245-1251.

		Reporting Item	Page Number
Title			1
	#1	Concise description of the nature and topic of the study identifying the study as qualitative or indicating the approach (e.g. ethnography, grounded theory) or data collection methods (e.g. interview, focus group) is recommended	
Abstract			2
	#2	Summary of the key elements of the study using the abstract format of the intended publication; typically includes background, purpose, methods, results and conclusions	
Introduction			4/5
Problem formulation	#3	Description and signifcance of the problem / phenomenon studied: review of relevant theory and empirical work; problem statement	4/5
Purpose or research question	#4	Purpose of the study and specific objectives or questions	4/5

Methods			6/7
Qualitative approach and research paradigm	#5	Qualitative approach (e.g. ethnography, grounded theory, case study, phenomenolgy, narrative research) and guiding theory if appropriate; identifying the research paradigm (e.g. postpositivist, constructivist / interpretivist) is also recommended; rationale. The rationale should briefly discuss the justification for choosing that theory, approach, method or technique rather than other options available; the assumptions and limitations implicit in those choices and how those choices influence study conclusions and transferability. As appropriate the rationale for several items might be discussed together.	6/7
Researcher characteristics and reflexivity	#6	Researchers' characteristics that may influence the research, including personal attributes, qualifications / experience, relationship with participants, assumptions and / or presuppositions; potential or actual interaction between researchers' characteristics and the research questions, approach, methods, results and / or transferability	1
Context	#7	Setting / site and salient contextual factors; rationale	
Sampling strategy	#8	How and why research participants, documents, or events were selected; criteria for deciding when no further sampling was necessary (e.g. sampling saturation); rationale	6
Ethical issues pertaining to human subjects	#9	Documentation of approval by an appropriate ethics review board and participant consent, or explanation for lack thereof; other confidentiality and data security issues	6/7
Data collection methods	#10	Types of data collected; details of data collection procedures including (as appropriate) start and stop dates of data collection and analysis, iterative process, triangulation of sources / methods, and modification of procedures in response to evolving study findings; rationale	5/6
Data collection instruments and technologies	#11	Description of instruments (e.g. interview guides, questionnaires) and devices (e.g. audio recorders) used for data collection; if / how the instruments(s) changed over the course of the study	5/6

Units of study	#12	Number and relevant characteristics of participants, documents, or events included in the study; level of participation (could be reported in results)	5
Data processing	#13	Methods for processing data prior to and during analysis, including transcription, data entry, data management and security, verification of data integrity, data coding, and anonymisation / deidentification of excerpts	5/6
Data analysis	#14	Process by which inferences, themes, etc. were identified and developed, including the researchers involved in data analysis; usually references a specific paradigm or approach; rationale	5/6
Techniques to enhance trustworthiness	#15	Techniques to enhance trustworthiness and credibility of data analysis (e.g. member checking, audit trail, triangulation); rationale	5/6
Results/findings			
Syntheses and interpretation	#16	Main findings (e.g. interpretations, inferences, and themes); might include development of a theory or model, or integration with prior research or theory	8
Links to empirical data	#17	Evidence (e.g. quotes, field notes, text excerpts, photographs) to substantiate analytic findings	8
Discussion			9
Intergration with prior work, implications, transferability and contribution(s) to the field	#18	Short summary of main findings; explanation of how findings and conclusions connect to, support, elaborate on, or challenge conclusions of earlier scholarship; discussion of scope of application / generalizability; identification of unique contributions(s) to scholarship in a discipline or field	
Limitations	#19	Trustworthiness and limitations of findings	9/10
Other			
Conflicts of interest	#20	Potential sources of influence of perceived influence on study conduct and conclusions; how these were managed	9/10
Funding	#21	Sources of funding and other support; role of funders in data collection, interpretation and reporting	1

None The SRQR checklist is distributed with permission of Wolters Kluwer © 2014 by the Association of American Medical Colleges. This checklist can be completed online using https://www.goodreports.org/, a tool made by the EQUATOR Network in collaboration with Penelope.ai

BMJ Open

Does Jenkins Sleep Scale show sex-related differential item functioning? Prospective cohort study among employees in public sector.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-074867.R1
Article Type:	Original research
Date Submitted by the Author:	03-Jan-2024
Complete List of Authors:	Juhola, Juhani; TYKS Turku University Hospital, Physical and Rehabilitation medicine Arokoski, J. P. A.; Helsingin yliopistollinen Keskussairaala, Department of Physical and Rehabilitation Medicine; Helsingin yliopisto, Ervasti, Jenni; Finnish Institute of Occupational Health, Topeliuksenkatu 41 a A, FI-00250, Helsinki, Finland, Kivimäki, Mika; University of Helsinki Faculty of Medicine, ; University College London, Dep. of Epidemiology and Public Health Vahtera, Jussi; Turun Yliopisto, Dep Public Health; Finnish Institute of Occupational Health, Myllyntausta, Saana; University of Turku, Saltychev, Mikhail; Turku University Hospital and University ofTurku, Turku, Finland , Department of Physical and Rehabilitation Medicine
Primary Subject Heading:	Public health
Secondary Subject Heading:	Research methods, Occupational and environmental medicine
Keywords:	SLEEP MEDICINE, Psychometrics, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Does Jenkins Sleep Scale show sex-related differential item functioning? Prospective cohort study among employees in public sector.

Running head: DIF of Jenkins Sleep Scale

Juhani Juhola MD[1], Jari Arokoski MD, PhD[2], Jenni Ervasti[3], Mika Kivimäki PhD[3,4,5], Jussi Vahtera PhD, MD[6,7], Saana Myllyntausta[8], Mikhail Saltychev MD, PhD[1]

[1] Department of Physical and Rehabilitation Medicine, Turku University Hospital and University of Turku, Turku, Finland

[2] Department of Physical and Rehabilitation Medicine, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

[3] Finnish Institute of Occupational Health, Helsinki, Finland

[4] Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[5] Department of Epidemiology and Public Health, University College London Medical School, London, United Kingdom

[6] Department of Public Health, University of Turku, Turku, Finland

[7] Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland

[8] Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland

Address for correspondence:

Juhani Juhola, Department of Physical and Rehabilitation Medicine, Turku University Hospital, PO Box 52, FI-20521, Turku, Finland

E-mail: juhani.juhola@gmail.com, Tel.: +358 40 738 9494, Fax: +358 2 313 3730

Keywords: sleep disorder; psychometrics; validity

ABSTRACT

Objectives:

To investigate if Jenkins Sleep Scale (JSS) demonstrates sex-related differential item functioning (DIF).

Design:

Prospective cohort study.

Setting:

Survey conducted by a national institute of occupational health.

Participants:

77,967 employees in public sector, mean age 51.9 (SD 13.1) years, 82% women.

Primary and secondary outcome measures:

Item response theory estimates: difficulty and discrimination parameters and differences in these parameters between males and females.

Results:

The mean JSS total score was 6.4 (4.8) points. For all four items, difficulty parameter demonstrated a slight shift towards underestimation the severity of sleep difficulties. The discrimination ability of all four items were moderate to high. For the JSS composite score, overall discrimination ability was moderate 0.98 (95% CI 0.97 to 0.99). There was slight uniform differential item functioning ($p<0.001$): two items showed better discrimination in men, while two other – in women.

Conclusions:

Among healthy population of employees in public sector, the JSS showed overall good psychometric properties. The JSS was able to discriminate people with different severity of sleep disturbances. However, when using the JSS, the respondents might slightly underestimate the severity of these disturbances. While the JSS may produce slightly different results when applied to men and women, these sex-related differences are probably neglectable when applied to clinical situations.

Strengths and limitations of this study

- The study was executed on a large cohort of almost 80,000 respondents employing sophisticated methods of item response theory
- The studied cohort was predominated by women
- The mean age was around 50 years and the results might be different in younger employees or during retirement transition
- The response rate was 70% without possibility to analyze missing responses

INTRODUCTION

There are numerous scales to assess the severity of sleep difficulties(1). Patient-reported outcome measures (PROMs) are easy to use and cost-efficient means to detect and grade sleep disturbances. The Jenkins Sleep Scale (JSS) has been developed as a brief and standardized questionnaire for sleep disturbances in 1998(2). It is one of the most used questionnaires in epidemiological studies(1-4). The JSS has been translated into several languages(5-10) and its psychometric properties have been found to be both valid and reliable across different patient groups, such as patients with rheumatoid(7) and psoriatic arthritis (6) and ankylosing spondylitis(5), fibromyalgia(10, 11), chest pain(12), and post cardiac surgery patients(2). Only a few studies have evaluated the psychometric properties of the JSS in large non-clinical populations(2, 3, 8, 13, 14).

Previous studies have found the JSS to be internally consistent amongst patients with fibromyalgia(10, 11), rheumatoid arthritis(7), ankylosing spondylitis(5) and psoriatic arthritis(6), as indicated by a Cronbach's alpha ranging from 0.7 to 0.9. Several studies have assessed the internal consistency of the JSS in a general population reporting respectively good to excellent Cronbach's alpha between 0.8 and 0.9(2, 3, 8, 9, 13, 14). A few previous studies have assessed the factor structure of the JSS and observed the JSS to be a unidimensional scale(3, 8, 9). The construct structure of the JSS analysis has also been assessed by a confirmatory factor analysis, showing strong correlations between all four items and a common factor(3). So far, no studies have focused on the psychometric properties of the JSS assessed by applying item response theory or Rasch analysis. Item response theory investigates the relationship between the performance on a test item and the average (in a particular population of interest) level of the ability that item was designed to measure. It does not assume that each item is equally difficult, where difficulty is understood as the level of measurable ability needed to get a particular response to an item. This differentiates item response theory from other methods, which assume equality of response difficulties when several items are measured on an ordinal scale. The item response theory suggests that these differences between item difficulties may clinically be relevant and they should be taken into account when interpreting the results obtained from a test with multiple items. Additionally, item response theory suggests that individual items, as well as an entire test, may perform differently in at the different levels of measurable ability.

Sex-related differences of sleep and circadian rhythm are well known(15, 16). It has been suggested that such differences may be age-related and occur as early as the middle years of life(15). Even though women may have better sleep quality than men in relation to sleep length, sleep-onset latency and sleep efficiency(17), women have 1.5 times higher risk to develop insomnia than men, and this predisposition has been found to be consistent and progressive with ageing(18, 19). Shorter circadian cycle lengths as well as a larger amplitude of circadian variation in women may lead to more frequent nighttime impairment in women(16). Sex differences in the sleep disorders underscore the need to account for sex

in sleep medicine and sleep research(16). Additionally, the diagnostics of diseases related to sleep disorders may differ between men and women. For example, narcolepsy or sleep apnea may be diagnosed later (or even remained undiagnosed) in women, at least partially due to variation in presenting symptoms(16, 20). Restless legs syndrome is also more common among women(16, 17, 21). It has also been suggested that the decreased need for sleep is associated with ageing (shorter sleep duration and nighttime awakenings) may be more common among men than women(22). Sex differences in the incidence of insomnia are the result of a complex combination of biopsychosocial factors changing across the life span(16-18, 20-26). These differences may be related to sex hormones or specific sex-dependent patterns of physiological periods like puberty, menstruation, pregnancy, and menopause or other causes(16-18, 20-26). Most of previous studies on the topic have focused on sex- and age-related differences in prevalence and incidence of insomnia, while milder sleep disturbances have less been studied.

Previous studies have suggested that both sex differences in sleep and circadian rhythms may impact evaluation of sleep disorders(16). Sleep scales, including the JSS, may possibly perform differently across sexes(24). For example, it has been reported that women may perceive nighttime awakenings as more difficult than men(22). The potential sex-related differential item functioning (DIF) of the JSS has not been studied before. The aim of this study was to investigate the psychometric properties of the JSS focusing especially on the potential sex-related DIF by applying the item response theory.

METHODS

The Finnish Public Sector (FPS) study is an ongoing prospective cohort study. The Finnish Public Sector study monitors the wellbeing and health of personnel in the municipal and wellbeing services sectors as well as their work and changes in it. The questionnaires of the studies are sent to over 100,000 employees every other year. The study covers six cities with more than 100,000 inhabitants and their five neighbouring cities as well as eight hospital districts. The study produces data of national significance. It is the country's largest and longest-standing survey of municipal and well-being sector personnel and it covers nearly 30% of municipal and well-being sector employees. The FPS study was originally started in 1997 and is officially called “Psychosocial Factors and Health: The Finnish Public Sector (FPS) study”. The study has been approved by the HUS Regional Committee on Medical Research Ethics (HUS/1210/2016). The ethical statement was last updated on 16 March 2023 due to the new research organisations created by the health and social services reform. There is no explicit informed consent form, but each respondent gets acquainted with the “Notice for the Kunta10 participants” (www.ttl.fi/en/tutkimus/hankkeet/kunta-ja-hyvinvointialan-henkiloston-seurantatutkimus-fps/kunta10-tiedote-tutkittavalle). When starting a survey, the respondents are aware that the survey results are utilised in scientific research. The response rate varies between the survey waves, usually staying around 70%.

All the data were obtained from the survey responses. Age was defined in full years at the time of survey response. Body mass index (BMI) was defined as weight divided by height to the power of two. The level of physical activity was calculated from the survey responses and converted into metabolic equivalent of task per hour per week (MET-h/week). Alcohol consumption was obtained from the survey and converted into gram/week. The respondents were asked about their usual amount of sleep hours per 24 hours with the following nine response alternatives: ≤6 hours, 6.5 hours, 7 hours, 7.5 hours, 8 hours, 8.5 hours, 9 hours, 9.5 hours, and ≥10 hours. The responses were then dichotomized as ≤7 hours vs. >7 hours of sleep.

The JSS is a four-item questionnaire used to grade the frequency of common sleep problems during the previous month(2): trouble falling asleep, waking up but no trouble falling asleep again, waking up and trouble falling asleep again, waking up feeling tired (i.e., waking up after the usual amount of sleep feeling tired and worn out). Each item was rated on a Likert-like scale from zero to five, where zero is “never”, one is “1-3 days”, two is “about 1 night/week”, three is “2-4 nights/week”, four is “5-6 nights/week” and five is “almost every night”. The total score is a simple sum of all four items’ scores and ranges from zero (“no sleep problems”) to 20 (“most sleep problems”). A score =<11 was considered as “little or no sleep disturbances” and >11 was considered as “high frequency of sleep disturbances”(27).

Statistical analysis

The results are reported as absolute numbers and percentages or as means and standard deviations (SD). The results were accompanied by 95% confidence intervals (95% CI) or two-tailed p-values, when appropriate. Using item response theory (IRT), the average level of reported sleep problems in the studied population was estimated based on the principle of maximum likelihood. Then, the level of sleep problems reported by each participant was compared to the average level observed in the entire sample. After fitting the model, both parameters – 'difficulty' and 'discrimination' – were calculated for each of the four items of the JSS by using the graded response model (GRM). Difficulty is the level of reported sleep problems needed to choose a particular response. In turn, discrimination is the steepness of the regression curve, with the severity of sleep problems placed on the X-axis and the expected score of the JSS on the Y-axis. Ideally, the steepest interval should correspond to the patients who obtained an average score of two or three. If such is the case, then a test (or an item) is especially sensitive in distinguishing people with a level of sleep problems below average from those with levels above average. In this study, discrimination of 0.01 to 0.34 was considered 'none' (a totally level regression curve) or 'very low'; 0.35 to 0.64 was considered 'low'; 0.65 to 1.34 was considered 'moderate'; 1.35 to 1.69 was considered 'high'; and a discrimination >1.7 was considered 'perfect' (a regression curve approaching a vertical line)(28). An item information curve helps to comprehend this graphically, appointing the steepest interval of the curve to the level of disability that is associated with the most information that can be obtained from the item. Item information is calculated as an invert standard error. Results were reported along with their 95% confidence intervals (95%CI). The item characteristic curves of all four items are available from the corresponding author on request.

Differential item functioning (DIF) is a statistical characteristic of a scale item (here counted for each of four items included in JSS) that describes if the item is measuring an ability (here severity of sleep problems) differently for separate subgroups (here sexes) within the sample. To assess a DIF, the probit logistic regression was used to test whether an item exhibits either uniform or nonuniform DIF between sex groups, that is, whether an item favors one group over the other for all values of severity of sleep problems or for only some values of that(29, 30). A uniform DIF occurs when the difference between groups remains the same across the entire scale. In turn, a nonuniform DIF is observed when the direction of difference between groups varies at different levels of sleep problems (e.g., if men perform better than women up to a midpoint and worse than women after that). A two-tailed p-value ≤0.05 indicated a significant difference between sexes. When a significant DIF was observed, the results of DIF analysis were also presented and evaluated graphically as item information function curves. An item information function describes the precision, which an item or the entire test achieves for different levels of sleep difficulties. To put it in a simpler way, an item information function is an inverse variance.

The analyses were performed using Stata/IC Statistical Software: Release 17. College Station (StataCorp LP, TX, USA).

RESULTS

Of 81,136 participants, 77,967 have responded to at least one item of the JSS. 14,349 (18%) were men and 63,618 (82%) were women (Table 1). Their mean age was 51.9 (SD 13.1) years, body mass index 26.2 (SD 4.7) kg/m^2, physical activity 29.6 (25.3) METs/week, and alcohol consumption was 50.1 (SD 91.3) g/week (equivalent to around four units of alcohol per week). Of the respondents, 56,014 (72%) were sleeping seven or less hours per night. The mean JSS total score was 6.4 (SD 4.8) points. Of the respondents, 12,629 (16%) had JSS total score more than 11.

Difficulty parameter of JSS

Table 2 shows the estimates of difficulty parameter for all four items of the JSS. All four items demonstrated a slight shift towards higher severity of sleep difficulties – the estimates close to zero could be seen at the lowest end (instead of the middle point) of the scale. In other words, the respondents tended to underestimate their sleep difficulties. For example, for the item ”trouble falling asleep”, the respondents with slightly worse than average sleep difficulties still tended to mark the minimal possible score of one point. This shift towards underestimation was, however, mild. The same mild shift toward underestimation of the sleep problems was seen for both sexes (Table 3 and Figure 1).

Discrimination parameter of JSS

The discrimination estimates for the item “Waking up and trouble falling asleep again” were high for both sexes: 1.92 for men and 2.04 for women (Table 3). For the other three items, the estimates were moderate ranging in both sexes from 0.71 to 1.16. The overall discrimination of the composite JSS score was moderate 0.98 (95% CI 0.97 to 0.99).

Differential item functioning (DIF) of JSS

When considering both discrimination and difficulty parameters, there were significant differences between sexes, $p<0.001$. Figure 1 shows a test characteristic curve for the entire cohort. Figure 2 presents the item information functions of each item grouped by sex. For every JSS item and for both sexes, the most information could be observed at the slightly elevated levels of sleep disturbances. As shown in Figure 2, the discrimination parameter was steeper in men for the JSS items ‘Trouble falling asleep’ and ‘Waking up feeling tired’. Respectively, the discrimination was steeper in women for the items ‘Waking up but no trouble falling sleep again’ and ’Waking up and trouble falling asleep again’. The shapes of curves were close to uniform for all the items.

DISCUSSION

In this survey-based cross-sectional cohort study among 77,967 employees of the public sector, there were little differences in the psychometric properties of JSS between sexes. All four items demonstrated a slight shift towards higher severity of sleep difficulties; the respondents tended (but only mildly) to underestimate their sleep difficulties. This shift was seen for both sexes. The discrimination estimates ranged from moderate to high, which means that the JSS is a sensitive scale for distinguishing people with different levels of sleep difficulties. A uniform DIF (slight but statistically significant) was present for all four items; the JSS was more sensitive among men for items ”Trouble falling asleep” and ”Waking up feeling tired” and among women for items ”Waking up but no trouble falling asleep again” and ”Waking up and trouble falling asleep again”. These differences may be related to different sleep disorders and to the differences of the incidence of these disorders between men and women. Women have more sex hormone related sleep disorders(16-18, 20-26) and also restless leg syndrome(17), while men have more obstructive sleep apnea and sleep disordered breathing, which are known to cause trouble falling asleep but also increasing daytime tiredness(16, 20)**.** These sex-related differences are that undeniable that although the behavioral treatments for insomnia have equal effects regardless of gender, certain pharmacologic treatments require different dosing based on sex(16).

The generalizability of the results might be weakened by the sex disbalance of the studied cohort (women were predominated) as fewer men work in the public sector in Finland. However, with almost 15 000 men in our data, it is unlikely that this is a source of a major bias. Also, the mean age of study participants was 52 years and, therefore, the results describe principally people in the last third of their working life span. While it has been widely used for over two decades, the Finnish translation of JSS had never undergone a full linguistic validation process, which might affect its equivalency with an English version. The response rate was 70% and there was no analysis of whether the non-respondents’ demographic characteristics might affect the results.

The direct comparison between the present results and previous research is limited by the fact that no earlier studies have focused on the psychometric properties of the JSS assessed by applying item response theory or Rasch analysis. This might leave the following clinically relevant questions unanswered: does a Likert-like scale used by the JSS behave similarly for all four items, does the JSS (as an entire test and its individual items) perform differently across the whole severity spectrum of sleep disturbances and does the JSS perform equally well in diverse subgroups and situations? As such, this also was the first study exploring the DIF of the JSS. However, the results of this study reflect previously observed differences in the amount and severity of sleep difficulties among men and women(16, 17, 20-26). The results are also in line with previously reported differences in the way men and women grade their sleep difficulties when responding to questionnaires(16, 24). Previous studies have suggested that both sex and gender differences in sleep and circadian rhythms may impact evaluation of sleep disorders and some sleep

scales, including the JSS, may behave differently among men and women(16). For example, the Pittsburgh Sleep Quality Index has showed similar sex-related inconsistency(31). The DIF has been detected for the Karolinska Sleep Questionnaire(32). Also, the Patient-Reported Outcomes Measurement Information System (PROMIS), which is a very popular standard general patient-reported outcome measure, has demonstrated an age-related DIF for a sleep- item(33).

The sex-related DIF of patient-reported outcome measures is a common finding. For example, such a DIF has been found for scales measuring quality of life, depression, disability caused by pain and general disability(34-40)

Simply explained, the significance of the results from a clinical point of view is that the JSS works almost equally well across both sexes. The DIF observed here was small and uniform, hardly affecting the interpretation or comparison of the scores obtained from a few individual people. On the other hand, this DIF may be of significant importance when the JSS is used to collect data from large populations, especially when comparing populations with dissimilar sex distributions. If there is such a situation, then that comparison should separately be performed by sex groups. This can be particularly true when, in addition to a composite score, research question concerns scores obtained from the JSS individual items.

Further research may reveal the potential DIF of the JSS among people of different age groups working in other fields than public sector, assuming that diverse physical and psychological work demands might affect the results obtained by the JSS. In addition, populations with different comorbidities (e.g., sleep apnea and disordered breathing or cardiovascular and metabolic disorders) may show results, which are different to the present ones.

Conclusions

The JSS showed overall good psychometric abilities, such as difficulty and discrimination, among public sector employees. The JSS was able to discriminate people with different severity of sleep disturbances. However, when using the JSS, the respondents might slightly underestimate the severity of these disturbances. Also, the JSS may produce slightly different results when applied to men or women. Nevertheless, even though these sex-related differences were statistically significant, they are probably neglectable when applied to clinical situations.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

TABLES AND FIGURES

Table 1. Descriptive characteristics of the study cohort

Variable	Men		Women		Entire cohort	
	N	%	N	%	N	%
Total					77,967	100
Sex						
Men	100	0	0	100	14,349	18
Women	0	100	100	0	63,618	82
Sleep						
≤7 hours/night	10,779	75	45,235	71	56,014	72
>7 hours/night	3,570	25	18,383	29	21,953	28
	Mean	SD	Mean	SD	Mean	SD
Age, years	52.5	13.3	51.7	13.1	51.9	13.1
Physical activity, MET-hour/week	33.3	29.9	28.8	24.1	29.6	25.3
Body mass index, kg/m^2	26.9	4.0	26.1	4.8	26.2	4.7
Alcohol consumption, grams	99.4	156.0	38.9	63.7	50.1	91.3
Jenkin's Sleep Scale, points	5.8	4.6	6.6	4.8	6.4	4.8

Table 2. Difficulty coefficients of the JSS items in both sexes together (n=77,967)

Items and scores	Difficulty	95% CI	
Trouble falling asleep			
1 vs. 0	0.18	0.16	0.19
2 vs. 1	1.19	1.18	1.21
3 vs. 2	0.85	0.83	0.86
4 vs. 3	2.66	2.64	2.69
5 vs. 4	1.31	1.28	1.33
Waking up but no trouble falling asleep again			
1 vs. 0	-0.79	-0.81	-0.78
2 vs. 1	0.22	0.20	0.24
3 vs. 2	-0.13	-0.14	-0.11
4 vs. 3	1.69	1.67	1.72
5 vs. 4	0.33	0.31	0.36
Waking up and trouble falling asleep again			
1 vs. 0	-0.23	-0.25	-0.22
2 vs. 1	0.78	0.77	0.80
3 vs. 2	0.44	0.42	0.45
4 vs. 3	2.25	2.23	2.28
5 vs. 4	0.90	0.87	0.92
Waking up feeling tired			
1 vs. 0	-0.54	-0.55	-0.52
2 vs. 1	0.48	0.46	0.50
3 vs. 2	0.13	0.12	0.15
4 vs. 3	1.95	1.93	1.98
5 vs. 4	0.59	0.57	0.62

Table 3. Difficulty and discrimination coefficients of the JSS items by sex

Items and scores	Men			Women		
	Coefficient	95% CI		Coefficient	95% CI	
Trouble falling asleep						
Discrimination	0.95	0.91	0.99	0.85	0.82	0.87
Difficulty						
1 vs. 0	0.41	0.36	0.47	0.42	0.39	0.45
2 vs. 1	1.22	1.15	1.29	1.29	1.25	1.34
3 vs. 2	1.26	1.17	1.34	1.11	1.07	1.16
4 vs. 3	3.05	2.87	3.23	3.24	3.14	3.34
5 vs. 4	1.03	0.84	1.23	1.04	0.95	1.13
Waking up but no trouble falling asleep again						
Discrimination	0.99	0.94	1.04	1.17	1.13	1.21
Difficulty						
1 vs. 0	-0.69	-0.75	-0.64	-0.76	-0.80	-0.72
2 vs. 1	0.60	0.53	0.67	0.60	0.56	0.63
3 vs. 2	0.05	-0.02	0.11	0.00	-0.03	0.03
4 vs. 3	1.84	1.74	1.94	1.76	1.70	1.81
5 vs. 4	0.21	0.11	0.31	0.35	0.31	0.40
Waking up and trouble falling asleep again						
Discrimination	1.92	1.81	2.04	2.04	1.96	2.11
Difficulty						
1 vs. 0	-0.10	-0.14	-0.07	0.04	0.02	0.07
2 vs. 1	0.58	0.54	0.61	0.67	0.64	0.70
3 vs. 2	0.62	0.58	0.66	0.66	0.63	0.68
4 vs. 3	1.65	1.59	1.72	1.70	1.65	1.74
5 vs. 4	1.34	1.27	1.40	1.38	1.34	1.42
Waking up feeling tired						
Discrimination	0.84	0.80	0.84	0.72	0.70	0.74
Difficulty						
1 vs. 0	-0.50	-0.56	-0.50	-0.53	-0.57	-0.49
2 vs. 1	0.61	0.55	0.61	0.73	0.69	0.77
3 vs. 2	0.43	0.36	0.43	0.20	0.16	0.24
4 vs. 3	2.26	2.14	2.26	2.40	2.33	2.48
5 vs. 4	1.06	0.93	1.06	0.94	0.88	1.01

Figure 1. Test characteristic curve in both sexes together

Figure 2. Item information functions of JSS items grouped by sex

Contributors All the authors (JJ, SM, JPAA, JE, MS, MK and JV) substantially contributed to the conception and design of the work, drafting the work and revised it critically for important intellectual content, interpreted the data, and finally approved the version published. All the authors achieved an agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. JJ was responsible for preparing the first draft. MS was responsible for the main data analysis. JV and MK were responsible for the acquisition of data. JV was a guarantor.

Funding This study was supported by funding granted by the Academy of Finland (grant 633666 to MK); NordForsk (to MK and JV); the UK MRC (grant K013351 to MK) and The Finnish Environment Fund (grants 190172 and 118060 to SM).

Competing interests None.

Patient consent for publication Not required.

Ethics approval The ethics committee of the Hospital District of Helsinki and Uusimaa has approved the study.

Data availability statement Data are available upon reasonable request. Individual-level survey data cannot be made publicly available, but information on the data and analyses is available upon request to the corresponding author.

References

1. Lallukka T, Dregan A, Armstrong D. Comparison of a sleep item from the General Health Questionnaire-12 with the Jenkins Sleep Questionnaire as measures of sleep disturbance. J Epidemiol. 2011;21(6):474-80.
2. Jenkins CD, Stanton BA, Niemcryk SJ, Rose RM. A scale for the estimation of sleep problems in clinical research. J Clin Epidemiol. 1988;41(4):313-21.
3. Tibubos AN, Zenger M, Schmalbach B, Beutel ME, Brahler E. Measurement invariance, validation and normative data of the Jenkins Sleep Scale-4 (JSS-4) in the German general population across the life span. J Psychosom Res. 2020;130:109933.
4. Vahtera J, Pentti J, Helenius H, Kivimaki M. Sleep disturbances as a predictor of long-term increase in sickness absence among employees after family death or illness. Sleep. 2006;29(5):673-82.
5. Duruoz MT, Ulutatar F, Ozturk EC, Unal-Ulutatar C, Sanal Toprak C, Kayhan O. Assessment of the validity and reliability of the Jenkins Sleep Scale in ankylosing spondylitis. Int J Rheum Dis. 2019;22(2):275-9.
6. Duruöz MT, Erdem D, Gencer K, Ulutatar F, Baklacıoğlu HŞ. Validity and reliability of the Turkish version of the Jenkins Sleep Scale in psoriatic arthritis. Rheumatology International. 2018;38:261-5.
7. Duruöz MT, Ünal Ç, Ulutatar F, Sanal Toprak C, Gündüz OH. The Validity and Reliability of Turkish Version of the Jenkins Sleep Evaluation Scale in Rheumatoid Arthritis. Arch Rheumatol. 2018;33(2):160-7.
8. Reis C, Mestre C, Tecedeiro MMV, Paiva T. Translation, cross-cultural adaptation and psychometric properties of the Jenkins Sleep Scale in a sample of Portuguese shift workers. Laboratório de Psicologia. 2014;12:89-98.
9. Salman M, Khan AH, Sulaiman SAS, Saleem F, Asif N, Mustafa ZU, et al. Psychometric validation of the Urdu version of the Jenkins sleep evaluation questionnaire. Journal of Pharmacy Practice and Community Medicine. 2018;4(1).
10. Unal-Ulutatar C, Ozsoy-Unubol T. Psychometric properties of Turkish version of Jenkins sleep scale in fibromyalgia syndrome. Advances in Rheumatology. 2020;60.

11. Crawford B, Piault E, Lai C, Sarzi-Puttini P. Assessing sleep in fibromyalgia: investigation of an alternative scoring method for the Jenkins Sleep Scale based on data from randomised controlled studies. Clinical and Experimental Rheumatology-Incl Supplements. 2010;28(6):S100.
12. Jerlock M, Gaston-Johansson F, Kjellgren KI, Welin C. Coping strategies, stress, physical activity and sleep in patients with unexplained chest pain. BMC nursing. 2006;5(1):1-10.
13. Nasermoaddeli A, Sekine M, Kumari M, Chandola T, Marmot M, Kagamimori S. Association of sleep quality and free time leisure activities in Japanese and British civil servants. Journal of occupational health. 2005;47(5):384-90.
14. Ornat L, Martínez-Dearth R, Chedraui P, Pérez-López FR. Assessment of subjective sleep disturbance and related factors during female mid-life with the Jenkins Sleep Scale. Maturitas. 2014;77(4):344-50.
15. Carrier J, Semba K, Deurveilher S, Drogos L, Cyr-Cronier J, Lord C, et al. Sex differences in age-related changes in the sleep-wake cycle. Frontiers in neuroendocrinology. 2017;47:66-85.
16. Marsella JL, Sharkey KM. Sex Differences in Sleep Disorders. In: Attarian H, Viola-Saltzman M, editors. Sleep Disorders in Women: A Guide to Practical Management. Cham: Springer International Publishing; 2020. p. 65-81.
17. Krishnan V, Collop NA. Gender differences in sleep disorders. Curr Opin Pulm Med. 2006;12(6):383-9.
18. Suh S, Cho N, Zhang J. Sex Differences in Insomnia: from Epidemiology and Etiology to Intervention. Curr Psychiatry Rep. 2018;20(9):69.
19. Zhang B, Wing Y-K. Sex Differences in Insomnia: A Meta-Analysis. Sleep. 2006;29(1):85-93.
20. Bublitz M, Adra N, Hijazi L, Shaib F, Attarian H, Bourjeily G. A Narrative Review of Sex and Gender Differences in Sleep Disordered Breathing: Gaps and Opportunities. Life. 2022;12(12):2003.
21. Meers J, Stout-Aguilar J, Nowakowski S. Chapter 3 - Sex differences in sleep health. In: Grandner MA, editor. Sleep and Health: Academic Press; 2019. p. 21-9.
22. Jonasdottir SS, Minor K, Lehmann S. Gender differences in nighttime sleep patterns and variability across the adult lifespan: a global-scale wearables study. Sleep. 2020;44(2).
23. Madrid-Valero JJ, Kirkpatrick RM, González-Javier F, Gregory AM, Ordoñana JR. Sex differences in sleep quality and psychological distress: Insights from a middle-aged twin sample from Spain. Journal of Sleep Research.n/a(n/a):e13714.
24. Mallampalli MP, Carter CL. Exploring sex and gender differences in sleep health: a Society for Women's Health Research Report. J Womens Health (Larchmt). 2014;23(7):553-62.
25. Mong JA, Baker FC, Mahoney MM, Paul KN, Schwartz MD, Semba K, et al. Sleep, rhythms, and the endocrine brain: influence of sex and gonadal hormones. J Neurosci. 2011;31(45):16107-16.
26. Mong JA, Cusmano DM. Sex differences in sleep: impact of biological sex and sex steroids. Philos Trans R Soc Lond B Biol Sci. 2016;371(1688):20150110.
27. Monterrosa-Castro Á, Portela-Buelvas K, Salguedo-Madrid M, Mo-Carrascal J, Leidy CD-M. Instruments to study sleep disorders in climacteric women. Sleep Science. 2016;9(3):169-78.
28. Baker FB. The basics of item response theory. Madison, WI, USA: ERIC Clearinghouse on Assessment and Evaluation; 2001.
29. De Boeck P, Wilson M. Explanatory item response models: A generalized linear and nonlinear approach: Springer; 2004.
30. Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. Journal of Educational measurement. 1990;27(4):361-70.
31. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. Sleep Med Rev. 2016;25:52-73.
32. Santos KD, Fischer FM, de Andrade DF, Barbetta PA, Alves Cardoso MR, de Gouveia Vilela RA. Shift work and sex are variants across Karolinska Sleep Questionnaire and the application of linking analysis as a step forward using Item Response Theory. Chronobiol Int. 2019;36(1):42-56.
33. van Kooten J, Terwee CB, Luijten MAJ, Steur LMH, Pillen S, Wolters NGJ, et al. Psychometric properties of the Patient-Reported Outcomes Measurement Information System (PROMIS) Sleep Disturbance and Sleep-Related Impairment item banks in adolescents. J Sleep Res. 2021;30(2):e13029.

34. Knott RJ, Lorgelly PK, Black N, Hollingsworth B. Differential item functioning in quality of life measurement: An analysis using anchoring vignettes. Soc Sci Med. 2017;190:247-55.
35. Saltychev M, Widbom-Kolhanen SS, Pernaa KI. Sex-related differential item functioning of neck disability index. Disabil Rehabil. 2023:1-7.
36. Katajapuu N, Laimi K, Heinonen A, Saltychev M. Floor and ceiling effects of the World Health Organization Disability Assessment Schedule 2.0 among patients with chronic musculoskeletal pain. Int J Rehabil Res. 2019;42(2):190-2.
37. Bourion-Bedes S, Schwan R, Laprevote V, Bedes A, Bonnet JL, Baumann C. Differential item functioning (DIF) of SF-12 and Q-LES-Q-SF items among french substance users. Health Qual Life Outcomes. 2015;13:172.
38. Cavanagh A, Wilson CJ, Caputi P, Kavanagh DJ. Symptom endorsement in men versus women with a diagnosis of depression: A differential item functioning approach. Int J Soc Psychiatry. 2016;62(6):549-59.
39. Verdam MGE, Oort FJ, Sprangers MAG. Item bias detection in the Hospital Anxiety and Depression Scale using structural equation modeling: comparison with other item bias detection methods. Qual Life Res. 2017;26(6):1439-50.
40. Brzoska P. Assessment of quality of life in individuals with chronic headache. Psychometric properties of the WHOQOL-BREF. BMC Neurol. 2020;20(1):267.

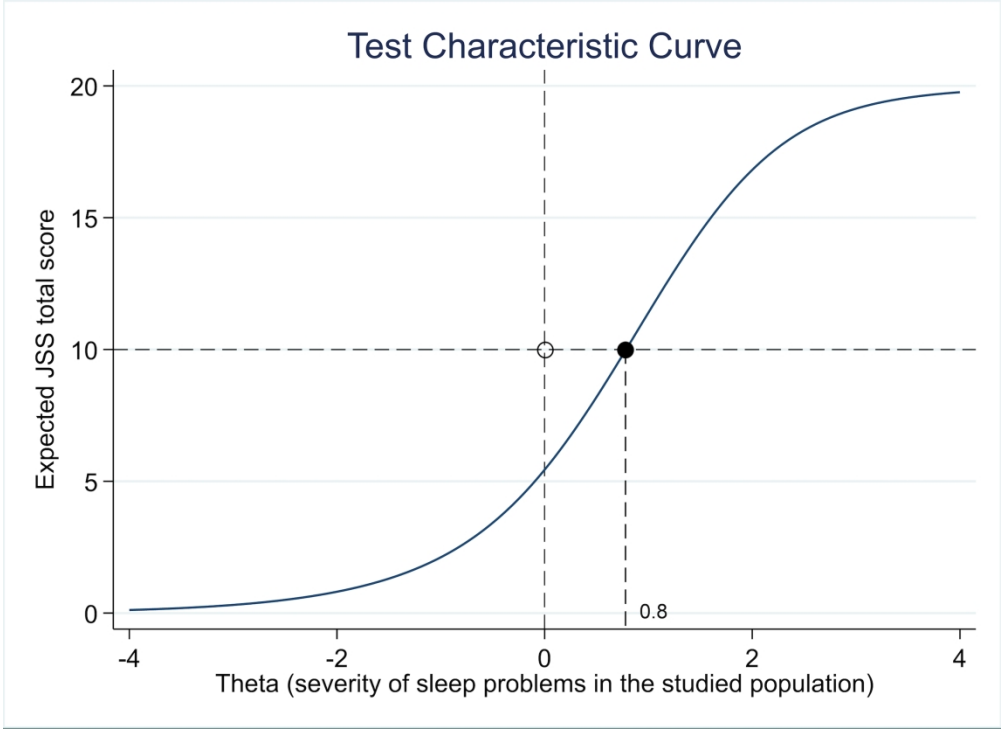

Figure 1. Test characteristic curve in both sexes together

87x63mm (762 x 762 DPI)

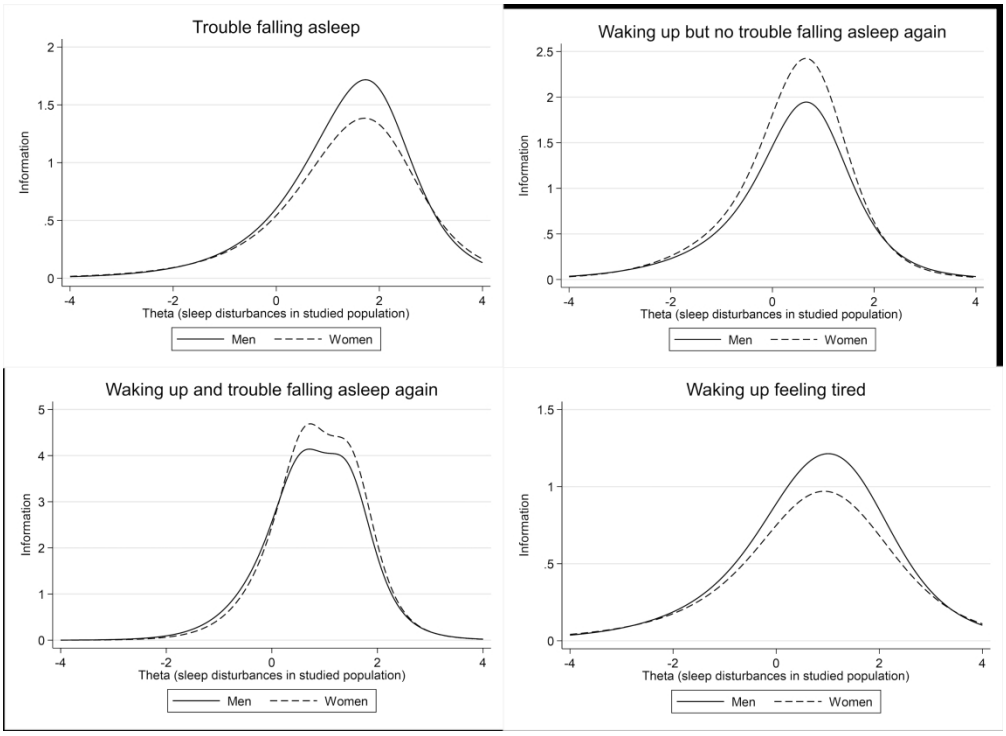

Figure 2. Item information functions of JSS items grouped by sex

147x107mm (762 x 762 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1-2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1-2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	3-4
Methods			
Study design	4	Present key elements of study design early in the paper	5-6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	5-6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5
Bias	9	Describe any efforts to address potential sources of bias	5-6
Study size	10	Explain how the study size was arrived at	5-6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5-6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5-6
		(*b*) Describe any methods used to examine subgroups and interactions	5-6
		(*c*) Explain how missing data were addressed	
		(*d*) If applicable, explain how loss to follow-up was addressed	5-6
		(*e*) Describe any sensitivity analyses	5-6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	5
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7
		(b) Indicate number of participants with missing data for each variable of interest	7
		(c) Summarise follow-up time (eg, average and total amount)	5
Outcome data	15*	Report numbers of outcome events or summary measures over time	7
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear	7

		which confounders were adjusted for and why they were included	
		(*b*) Report category boundaries when continuous variables were categorized	7
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	8
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	8-9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8-9
Generalisability	21	Discuss the generalisability (external validity) of the study results	9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	14

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ Open

Sex-related differential item functioning of the Jenkins Sleep Scale: a cross-sectional study among 77,967 employees in the Finnish public sector

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-074867.R2
Article Type:	Original research
Date Submitted by the Author:	27-Feb-2024
Complete List of Authors:	Juhola, Juhani; TYKS Turku University Hospital, Physical and Rehabilitation medicine Arokoski, J. P. A.; Helsingin yliopistollinen Keskussairaala, Department of Physical and Rehabilitation Medicine; Helsingin yliopisto, Ervasti, Jenni; Finnish Institute of Occupational Health, Topeliuksenkatu 41 a A, FI-00250, Helsinki, Finland, Kivimäki, Mika; University of Helsinki Faculty of Medicine, ; University College London, Dep. of Epidemiology and Public Health Vahtera, Jussi; Turun Yliopisto, Dep Public Health; Finnish Institute of Occupational Health, Myllyntausta, Saana; University of Turku, Saltychev, Mikhail; Turku University Hospital and University ofTurku, Turku, Finland , Department of Physical and Rehabilitation Medicine
Primary Subject Heading:	Public health
Secondary Subject Heading:	Research methods, Occupational and environmental medicine
Keywords:	SLEEP MEDICINE, Psychometrics, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Sex-related differential item functioning of the Jenkins Sleep Scale: a cross-sectional study among 77,967 employees in the Finnish public sector

Juhani Juhola MD[1], Jari Arokoski MD, PhD[2], Jenni Ervasti[3], Mika Kivimäki PhD[3,4,5], Jussi Vahtera PhD, MD[6,7], Saana Myllyntausta[8], Mikhail Saltychev MD, PhD[1]

[1] Department of Physical and Rehabilitation Medicine, Turku University Hospital and University of Turku, Turku, Finland

[2] Department of Physical and Rehabilitation Medicine, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

[3] Finnish Institute of Occupational Health, Helsinki, Finland

[4] Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[5] Department of Epidemiology and Public Health, University College London Medical School, London, United Kingdom

[6] Department of Public Health, University of Turku, Turku, Finland

[7] Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland

[8] Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland

Correspondence to:

Juhani Juhola, Department of Physical and Rehabilitation Medicine, Turku University Hospital, PO Box 52, FI-20521, Turku, Finland

E-mail: juhani.juhola@gmail.com

Tel.: +358 40 738 9494

Fax: +358 2 313 3730

KEYWORDS

sleep disorder; psychometrics; validity

ABSTRACT

Objectives

To investigate if the Jenkins Sleep Scale (JSS) demonstrates sex-related differential item functioning (DIF).

Design

Cross-sectional study.

Setting

Survey data from the Finnish Public Sector (FPS) study (2015-2017).

Participants

77,967 employees in Finnish public sector, mean age 51.9 (SD 13.1) years, 82% women.

Outcome measures

Item response theory estimates: difficulty and discrimination parameters of the JSS and differences in these parameters between males and females.

Results

The mean JSS total score was 6.4 (4.8) points. For all four items of the JSS, the difficulty parameter demonstrated a slight shift towards underestimation of the severity of sleep difficulties. The discrimination ability of all four items were moderate to high. For the JSS composite score, overall discrimination ability was moderate 0.98 (95% CI 0.97 to 0.99). Mild uniform differential item functioning ($p<0.001$) was seen: two items showed better discrimination ability among men, and two others among women.

Conclusions

The JSS showed overall good psychometric properties among this healthy population of employees in the Finnish public sector. The JSS was able to discriminate people with different severities of sleep disturbances. However, when using the JSS, the respondents might slightly underestimate the severity of these disturbances. While the JSS may produce slightly different results when answered by men and women, these sex-related differences are probably negligible when applied to clinical situations.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The study was executed on a large sample of almost 80,000 respondents, employing sophisticated methods of the item response theory.

- The studied sample was predominated by women.
- The mean age was around 50 years; and the results might be different among younger respondents or during retirement transition.
- The response rate of the surveys varied from 57% to 70%, without the possibility to analyse missing responses.

INTRODUCTION

There are numerous scales to assess the severity of sleep difficulties(1). Patient-reported outcome measures (PROMs) are easy to use and cost-efficient means to detect and grade sleep disturbances. The Jenkins Sleep Scale (JSS) has been developed as a brief and standardized questionnaire for sleep disturbances in 1998(2). It is one of the most used questionnaires in epidemiological studies(1-4). The JSS has been translated into several languages(5-10) and its psychometric properties have been found to be both valid and reliable across different patient groups, such as patients with rheumatoid(7) and psoriatic arthritis (6) and ankylosing spondylitis(5), fibromyalgia(10, 11), chest pain(12), and post cardiac surgery patients(2). Only a few studies have evaluated the psychometric properties of the JSS in large non-clinical populations(2, 3, 8, 13, 14).

Previous studies have found the JSS to be internally consistent among patients with fibromyalgia(10, 11), rheumatoid arthritis(7), ankylosing spondylitis(5) and psoriatic arthritis(6), as indicated by a Cronbach's alpha, which had been ranging from 0.7 to 0.9. Several studies have assessed the internal consistency of the JSS in a general population reporting respectively good to excellent Cronbach's alpha between 0.8 and 0.9(2, 3, 8, 9, 13, 14). A few previous studies have assessed the factor structure of the JSS and observed the JSS to be a unidimensional scale(3, 8, 9). The construct structure of the JSS has also been assessed by a confirmatory factor analysis, showing strong correlations between all four items and a common factor(3). So far, no studies have focused on the psychometric properties of the JSS by applying item response theory or Rasch analysis. Item response theory investigates the relationship between the performance of a test item and the average (in a particular population of interest) level of the ability that the item was designed to measure. It does not assume that each item is equally difficult, where difficulty is understood as the level of measurable ability needed to get a particular response to an item. This differentiates item response theory from other methods, which assume equality of response difficulties when several items are measured on an ordinal scale. The item response theory suggests that these differences between item difficulties may be clinically relevant and should be taken into account when interpreting the results obtained from a test with multiple items. Additionally, item response theory suggests that individual items, as well as an entire test, may perform differently at different levels of assessed ability.

Sex-related differences of sleep and circadian rhythm are well known(15, 16). It has been suggested that these differences may be age-related and might start at middle age(15). Even though women may have better sleep quality than men in relation to sleep length, sleep-onset latency and sleep efficiency(17), women have 1.5 times higher risk to develop insomnia than men, and this predisposition has been found to be consistent and progressive with ageing(18, 19). Shorter circadian cycle lengths as well as a larger amplitude of circadian variation in women may lead to more frequent nighttime impairment in women(16). Sex differences in the sleep disorders underscore the need to account for sex in sleep

medicine and sleep research(16). Additionally, the diagnostics of diseases related to sleep disorders may differ between men and women. For example, narcolepsy or sleep apnoea may be diagnosed later (or even remained undiagnosed) in women, at least partially due to variation in presenting symptoms(16, 20). Restless legs syndrome is more common among women(16, 17, 21). It has been suggested that the decreased need for sleep is associated with ageing (shorter sleep duration and nighttime awakenings) and may be more common among men than women(22). Sex differences in the incidence of insomnia are the result of a complex combination of biopsychosocial factors changing across the life span(16-18, 20-26). These differences may be related to hormones or specific sex-dependent patterns of physiological periods like puberty, menstruation, pregnancy, and menopause or other causes(16-18, 20-26). Most of previous studies on the topic have focused on sex- and age-related differences in prevalence and incidence of insomnia, while milder sleep disturbances have been less studied.

Previous studies have suggested that both sex differences in both sleep and circadian rhythms may impact evaluation of sleep disorders(16). Sleep scales, including the JSS, may possibly perform differently across sexes(24). For example, it has been reported that women may perceive nighttime awakenings more difficult than men(22). The potential sex-related differential item functioning (DIF) of the JSS has not been studied before. The aim of this study was to investigate the psychometric properties of the JSS focusing especially on the potential sex-related DIF by applying the item response theory.

METHODS

Study design

The Finnish Public Sector (FPS) study is an ongoing prospective study. The FPS survey data used in the present cross-sectional analysis were collected from the employees of the participating organizations in 2015 (hospitals) and in 2016 (municipalities). There were 76,760 employees eligible for these surveys, of which 53,505 (70%) responded. In addition, data were used from the 2017 survey sent to people that had left their employer by 2016 but had responded to at least one survey before that. There were 48,645 persons eligible for the 2017 survey, of which 27,631 (57%) responded. The study has been approved by the HUS Regional Committee on Medical Research Ethics (HUS/1210/2016). The ethical statement was last updated on March 16th, 2023, due to the new research organisations created by the health and social services reform. There is no explicit informed consent form, but each respondent was informed of the "Notice for the Kunta10 participants" (www.ttl.fi/en/tutkimus/hankkeet/kunta-ja-hyvinvointialan-henkiloston-seurantatutkimus-fps/kunta10-tiedote-tutkittavalle). When starting a survey, the respondents are aware that the survey results are utilised in scientific research.

All the data have been obtained from the survey responses. Age was defined in full years at the time of the survey response. Body mass index (BMI) was defined as weight divided by height to the power of two. The level of physical activity was calculated from the survey responses and converted into a metabolic equivalent of task per hour per week (MET-h/week). Alcohol consumption was obtained from the survey and converted into grams/week. The respondents were asked about their usual amount of sleep hours per 24 hours with the following nine response alternatives: ≤6 hours, 6.5 hours, 7 hours, 7.5 hours, 8 hours, 8.5 hours, 9 hours, 9.5 hours, and ≥10 hours. The responses were then dichotomized as ≤7 hours vs. >7 hours of sleep.

The JSS is a four-item questionnaire used to grade the frequency of common sleep problems during the previous month(2): trouble falling asleep, waking up but no trouble falling asleep again, waking up and trouble falling asleep again, waking up feeling tired (i.e., waking up after the usual amount of sleep feeling tired and worn out). Each item was rated on a Likert-like scale from zero to five, where zero is "never", one is "1-3 days", two is "about 1 night/week", three is "2-4 nights/week", four is "5-6 nights/week" and five is "almost every night". The total score is a simple sum of all four items' scores and ranges from zero ("no sleep problems") to 20 ("most sleep problems"). A score =<11 was considered as "little or no sleep disturbances" and >11 was considered as "high frequency of sleep disturbances"(27).

Statistical analysis

The results were reported as absolute numbers and percentages or as means and standard deviations (SD). The results were accompanied by 95% confidence intervals (95% CI) or two-tailed p-values, when appropriate. Using item response theory (IRT), the average level of reported sleep problems in the studied

population was estimated based on the principle of maximum likelihood. Then, the level of sleep problems reported by each participant were compared to the average level observed in the entire sample. After fitting the model, both parameters – 'difficulty' and 'discrimination' – were calculated for each of the four items of the JSS by using the graded response model (GRM). Difficulty is the level of reported sleep problems needed to choose a particular response. In turn, discrimination is the steepness of the regression curve, with the severity of sleep problems placed on the X-axis and the expected score of the JSS on the Y-axis. Ideally, the steepest interval should correspond to the patients who obtained an average score of two or three. If such is the case, then a test (or an item) is especially sensitive in distinguishing people with a level of sleep problems below average from those with levels above average. In this study, discrimination of 0.01 to 0.34 was considered 'none' (a completely level regression curve) or 'very low'; 0.35 to 0.64 was considered 'low'; 0.65 to 1.34 was considered 'moderate'; 1.35 to 1.69 was considered 'high'; and a discrimination >1.7 was considered 'perfect' (a regression curve approaching a vertical line)(28). An item information curve helps to comprehend this graphically, appointing the steepest interval of the curve to the level of disability that is associated with the most information that can be obtained from the item. Item information is calculated as an invert standard error. Results were reported along with their 95% confidence intervals (95%CI). The item characteristic curves of all four items are available from the corresponding author on request.

Differential item functioning (DIF) is a statistical characteristic of a scale item (here counted for each of four items included in JSS) that describes if the item is measuring an ability (here severity of sleep problems) differently for separate subgroups (here sexes) within the sample. To assess a DIF, the probit logistic regression was used to test whether an item exhibits either uniform or nonuniform DIF between sex groups, that is, whether an item favours one group over the other for all values of severity of sleep problems or for only some values(29, 30). A uniform DIF occurs when the difference between groups remains the same across the entire scale. In turn, a nonuniform DIF is observed when the direction of difference between groups varies at different levels of sleep problems (e.g., if men perform better than women up to a midpoint and worse than women after that). A two-tailed p-value ≤0.05 indicated a significant difference between sexes. When a significant DIF was observed, the results of DIF analysis were also presented and evaluated graphically as item information function curves. An item information function describes the precision, which an item or the entire test achieves for different levels of sleep difficulties. To put it in a simpler way, an item information function is an inverse variance.

The analyses were performed using Stata/IC Statistical Software: Release 17. College Station (StataCorp LP, TX, USA).

Patient and public involvement

None.

RESULTS

In total, there were 125,405 eligible participants in the 2015-2017 surveys. Of the respondents (n=81,136), all who answered to at least one JSS item were included for analysis (n=77,967). 14,349 (18%) were men and 63,618 (82%) were women (Table 1). Their mean age was 51.9 (SD 13.1) years, body mass index 26.2 (SD 4.7) kg/m^2, physical activity 29.6 (25.3) METs/week, and alcohol consumption was 50.1 (SD 91.3) g/week (equivalent to around four units of alcohol per week). Of the respondents, 56,014 (72%) were sleeping seven or less hours per night. The mean JSS total score was 6.4 (SD 4.8) points. Of the respondents, 12,629 (16%) had JSS total score more than 11.

Table 1. Descriptive characteristics of the study sample

Variable	Men		Women		Entire sample	
	N	%	N	%	N	%
Total					77,967	100
Sex						
Men	100	0	0	100	14,349	18
Women	0	100	100	0	63,618	82
Sleep						
≤7 hours/night	10,779	75	45,235	71	56,014	72
>7 hours/night	3,570	25	18,383	29	21,953	28
	Mean	SD	Mean	SD	Mean	SD
Age, years	52.5	13.3	51.7	13.1	51.9	13.1
Physical activity, MET-hour/week	33.3	29.9	28.8	24.1	29.6	25.3
Body mass index, kg/m^2	26.9	4.0	26.1	4.8	26.2	4.7
Alcohol consumption, grams	99.4	156.0	38.9	63.7	50.1	91.3
Jenkin's Sleep Scale, points	5.8	4.6	6.6	4.8	6.4	4.8

Difficulty parameter of JSS

Table 2 shows the estimates of the difficulty parameter for all four items of the JSS. All four items demonstrated a slight shift towards higher severity of sleep difficulties – the estimates close to zero could be seen at the lowest end (instead of the middle point) of the scale. In other words, the respondents tended to underestimate their sleep difficulties. For example, for the item "trouble falling asleep", the respondents with slightly worse than average sleep difficulties still tended to mark the minimal possible score of one point. This shift towards underestimation was, however, mild. The same mild shift toward underestimation of the sleep problems was seen for both sexes (Table 3 and Figure 1).

Table 2. Difficulty coefficients of the JSS items in both sexes together (n=77,967)

Items and scores	Difficulty	95% CI	
Trouble falling asleep			
1 vs. 0	0.18	0.16	0.19

2 vs. 1	1.19	1.18	1.21
3 vs. 2	0.85	0.83	0.86
4 vs. 3	2.66	2.64	2.69
5 vs. 4	1.31	1.28	1.33
Waking up but no trouble falling asleep again			
1 vs. 0	-0.79	-0.81	-0.78
2 vs. 1	0.22	0.20	0.24
3 vs. 2	-0.13	-0.14	-0.11
4 vs. 3	1.69	1.67	1.72
5 vs. 4	0.33	0.31	0.36
Waking up and trouble falling asleep again			
1 vs. 0	-0.23	-0.25	-0.22
2 vs. 1	0.78	0.77	0.80
3 vs. 2	0.44	0.42	0.45
4 vs. 3	2.25	2.23	2.28
5 vs. 4	0.90	0.87	0.92
Waking up feeling tired			
1 vs. 0	-0.54	-0.55	-0.52
2 vs. 1	0.48	0.46	0.50
3 vs. 2	0.13	0.12	0.15
4 vs. 3	1.95	1.93	1.98
5 vs. 4	0.59	0.57	0.62

Table 3. Difficulty and discrimination coefficients of the JSS items by sex

Items and scores	Men			Women		
	Coefficient	95% CI		Coefficient	95% CI	
Trouble falling asleep						
Discrimination	0.95	0.91	0.99	0.85	0.82	0.87
Difficulty						
1 vs. 0	0.41	0.36	0.47	0.42	0.39	0.45
2 vs. 1	1.22	1.15	1.29	1.29	1.25	1.34
3 vs. 2	1.26	1.17	1.34	1.11	1.07	1.16
4 vs. 3	3.05	2.87	3.23	3.24	3.14	3.34
5 vs. 4	1.03	0.84	1.23	1.04	0.95	1.13
Waking up but no trouble falling asleep again						
Discrimination	0.99	0.94	1.04	1.17	1.13	1.21
Difficulty						
1 vs. 0	-0.69	-0.75	-0.64	-0.76	-0.80	-0.72
2 vs. 1	0.60	0.53	0.67	0.60	0.56	0.63
3 vs. 2	0.05	-0.02	0.11	0.00	-0.03	0.03
4 vs. 3	1.84	1.74	1.94	1.76	1.70	1.81
5 vs. 4	0.21	0.11	0.31	0.35	0.31	0.40
Waking up and trouble falling asleep again						
Discrimination	1.92	1.81	2.04	2.04	1.96	2.11
Difficulty						
1 vs. 0	-0.10	-0.14	-0.07	0.04	0.02	0.07
2 vs. 1	0.58	0.54	0.61	0.67	0.64	0.70

3 vs. 2	0.62	0.58	0.66	0.66	0.63	0.68
4 vs. 3	1.65	1.59	1.72	1.70	1.65	1.74
5 vs. 4	1.34	1.27	1.40	1.38	1.34	1.42
Waking up feeling tired						
Discrimination	0.84	0.80	0.84	0.72	0.70	0.74
Difficulty						
1 vs. 0	-0.50	-0.56	-0.50	-0.53	-0.57	-0.49
2 vs. 1	0.61	0.55	0.61	0.73	0.69	0.77
3 vs. 2	0.43	0.36	0.43	0.20	0.16	0.24
4 vs. 3	2.26	2.14	2.26	2.40	2.33	2.48
5 vs. 4	1.06	0.93	1.06	0.94	0.88	1.01

Discrimination parameter of JSS

The discrimination estimates for the item "Waking up and trouble falling asleep again" were high for both sexes: 1.92 for men and 2.04 for women (Table 3). For the other three items, the estimates were moderate ranging in both sexes from 0.71 to 1.16. The overall discrimination of the composite JSS score was moderate 0.98 (95% CI 0.97 to 0.99).

Differential item functioning (DIF) of JSS

When considering both discrimination and difficulty parameters, there were significant differences between sexes, $p<0.001$. Figure 1 shows a test characteristic curve for the entire sample. Figure 2 presents the item information functions of each item grouped by sex. For every JSS item and for both sexes, the most information could be observed at the slightly elevated levels of sleep disturbances. As shown in Figure 2, the discrimination parameter was steeper for men for the JSS items 'Trouble falling asleep' and 'Waking up feeling tired'. Respectively, the discrimination was steeper for women for the items 'Waking up but no trouble falling sleep again' and 'Waking up and trouble falling asleep again'. The shapes of curves were close to uniform for all the items.

DISCUSSION

In this survey-based cross-sectional study among 77,967 employees in the Finnish public sector, there were minor differences in the psychometric properties of JSS between sexes. All four items demonstrated a slight shift towards higher severity of sleep difficulties; the respondents tended (but only mildly) to underestimate their sleep difficulties. This shift was seen for both sexes. The discrimination estimates ranged from moderate to high, which means that the JSS is a sensitive scale for distinguishing people with different levels of sleep difficulties. A uniform DIF (slight but statistically significant) was present for all four items; the JSS was more sensitive among men for items “Trouble falling asleep” and ”Waking up feeling tired” and among women for items ”Waking up but no trouble falling asleep again” and ”Waking up and trouble falling asleep again”. These differences may be related to different sleep disorders and to the differences of the incidence of these disorders between men and women. Women have more hormone related sleep disorders(16-18, 20-26) and also restless leg syndrome(17), while men have more obstructive sleep apnoea and breathing disorders related to sleep difficulties, which are known to cause trouble falling asleep but also increasing daytime tiredness(16, 20). While behavioural treatment of insomnia has equal effects for both sexes, some pharmacologic treatments may require different dosages based on sex(16).

The generalizability of the results might be weakened by the sex disbalance of the studied sample (women were predominated) as fewer men work in the public sector in Finland. However, with almost 15 000 men in our data, it is unlikely that this is a source of a major bias. Also, the mean age of study participants was 52 years and, therefore, the results describe principally people in the last third of their working life span. While it has been widely used for over two decades, the Finnish translation of JSS has never undergone a full linguistic validation process, which might affect its equivalency with an English version. The response rate in the surveys were 57% in 2015 – 2016 and 70% in 2017. No analyses were conducted on whether the demographic characteristics of non-respondents might affect the results.

The direct comparison between the present results and previous research is limited since no earlier studies have focused on the psychometric properties of the JSS applying item response theory or Rasch analysis. This might leave the following clinically relevant questions unanswered: does a Likert-like scale used by the JSS behave similarly for all four items, does the JSS (as an entire test and its individual items) perform differently across the whole severity spectrum of sleep disturbances and does the JSS perform equally well in diverse subgroups and situations? Moreover, this was also the first study exploring the DIF of the JSS. However, the results of this study reflect previously observed differences in the amount and severity of sleep difficulties among men and women(16, 17, 20-26). The results are also in line with previously reported differences in the way men and women grade their sleep difficulties when responding to questionnaires(16, 24). Previous studies have suggested that sex-related differences in sleep and circadian rhythms may affect the evaluation of sleep disorders by some scales, including the JSS(16). For

example, the Pittsburgh Sleep Quality Index has showed similar sex-related inconsistencies(31). The DIF has been reported for the Karolinska Sleep Questionnaire(32). Also, the Patient-Reported Outcomes Measurement Information System (PROMIS), a very popular standard general patient-reported outcome measure, has demonstrated an age-related DIF regarding sleep(33).

The sex-related DIF of patient-reported outcome measures is a common finding. For example, such a DIF has been found for scales measuring quality of life, depression, disability caused by pain and general disability(34-40)

The significance of the results from a clinical point of view is that the JSS performs relatively well for both sexes. The DIF observed here was minor and uniform, hardly affecting the practical interpretation of the JSS scores. On the other hand, this DIF may be of significant importance when the JSS is used to collect data from large populations, especially when comparing populations with dissimilar sex distributions. If there is such a situation, then that comparison should separately be performed by sex groups. This can be particularly true when, in addition to a composite score, research question concerns scores obtained from the JSS individual items.

Further research may reveal the potential DIF of the JSS among people of different age groups working in other fields than the public sector, assuming that diverse physical and psychological work demands might affect the results obtained by the JSS. In addition, populations with different comorbidities (e.g., sleep apnoea and disordered breathing or cardiovascular and metabolic disorders) may show results, which are different to the present ones.

CONCLUSIONS

The JSS showed overall good psychometric abilities, such as difficulty and discrimination, among public sector employees. The JSS was able to discriminate people with different severity of sleep disturbances. However, when using the JSS, the respondents might slightly underestimate the severity of these disturbances. Also, the JSS may produce slightly different results when applied to men or women. Nevertheless, even though these sex-related differences were statistically significant, they are probably negligible when applied to clinical situations.

CONTRIBUTORS

All the authors (JJ, SM, JPAA, JE, MS, MK and JV) substantially contributed to the conception and design of the work, drafting the work and revised it critically for important intellectual content, interpreted the data, and finally approved the version published. All the authors achieved an agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. JJ was responsible for preparing the first draft. MS was responsible for the main data analysis. JV and MK were responsible for the acquisition of data. JV was a guarantor.

FUNDING

This study was supported by funding granted by the Academy of Finland (grant 633666 to MK); NordForsk (to MK and JV); the UK MRC (grant K013351 to MK) and The Finnish Environment Fund (grants 190172 and 118060 to SM).

COMPETING INTERESTS

None.

PATIENT CONSENT FOR PUBLICATION

Not required.

ETHICS APPROVAL

The ethics committee of the Hospital District of Helsinki and Uusimaa has approved the study.

DATA AVAILABILITY STATEMENT

Data are available upon reasonable request. Individual-level survey data cannot be made publicly available, but information on the data and analyses is available upon request to the corresponding author.

REFERENCES

1. Lallukka T, Dregan A, Armstrong D. Comparison of a sleep item from the General Health Questionnaire-12 with the Jenkins Sleep Questionnaire as measures of sleep disturbance. J Epidemiol. 2011;21(6):474-80.
2. Jenkins CD, Stanton BA, Niemcryk SJ, Rose RM. A scale for the estimation of sleep problems in clinical research. J Clin Epidemiol. 1988;41(4):313-21.
3. Tibubos AN, Zenger M, Schmalbach B, Beutel ME, Brahler E. Measurement invariance, validation and normative data of the Jenkins Sleep Scale-4 (JSS-4) in the German general population across the life span. J Psychosom Res. 2020;130:109933.
4. Vahtera J, Pentti J, Helenius H, Kivimaki M. Sleep disturbances as a predictor of long-term increase in sickness absence among employees after family death or illness. Sleep. 2006;29(5):673-82.
5. Duruoz MT, Ulutatar F, Ozturk EC, Unal-Ulutatar C, Sanal Toprak C, Kayhan O. Assessment of the validity and reliability of the Jenkins Sleep Scale in ankylosing spondylitis. Int J Rheum Dis. 2019;22(2):275-9.
6. Duruöz MT, Erdem D, Gencer K, Ulutatar F, Baklacıoğlu HŞ. Validity and reliability of the Turkish version of the Jenkins Sleep Scale in psoriatic arthritis. Rheumatology International. 2018;38:261-5.
7. Duruöz MT, Ünal Ç, Ulutatar F, Sanal Toprak C, Gündüz OH. The Validity and Reliability of Turkish Version of the Jenkins Sleep Evaluation Scale in Rheumatoid Arthritis. Arch Rheumatol. 2018;33(2):160-7.
8. Reis C, Mestre C, Tecedeiro MMV, Paiva T. Translation, cross-cultural adaptation and psychometric properties of the Jenkins Sleep Scale in a sample of Portuguese shift workers. Laboratório de Psicologia. 2014;12:89-98.
9. Salman M, Khan AH, Sulaiman SAS, Saleem F, Asif N, Mustafa ZU, et al. Psychometric validation of the Urdu version of the Jenkins sleep evaluation questionnaire. Journal of Pharmacy Practice and Community Medicine. 2018;4(1).
10. Unal-Ulutatar C, Ozsoy-Unubol T. Psychometric properties of Turkish version of Jenkins sleep scale in fibromyalgia syndrome. Advances in Rheumatology. 2020;60.
11. Crawford B, Piault E, Lai C, Sarzi-Puttini P. Assessing sleep in fibromyalgia: investigation of an alternative scoring method for the Jenkins Sleep Scale based on data from randomised controlled studies. Clinical and Experimental Rheumatology-Incl Supplements. 2010;28(6):S100.
12. Jerlock M, Gaston-Johansson F, Kjellgren KI, Welin C. Coping strategies, stress, physical activity and sleep in patients with unexplained chest pain. BMC nursing. 2006;5(1):1-10.

13. Nasermoaddeli A, Sekine M, Kumari M, Chandola T, Marmot M, Kagamimori S. Association of sleep quality and free time leisure activities in Japanese and British civil servants. Journal of occupational health. 2005;47(5):384-90.

14. Ornat L, Martínez-Dearth R, Chedraui P, Pérez-López FR. Assessment of subjective sleep disturbance and related factors during female mid-life with the Jenkins Sleep Scale. Maturitas. 2014;77(4):344-50.

15. Carrier J, Semba K, Deurveilher S, Drogos L, Cyr-Cronier J, Lord C, et al. Sex differences in age-related changes in the sleep-wake cycle. Frontiers in neuroendocrinology. 2017;47:66-85.

16. Marsella JL, Sharkey KM. Sex Differences in Sleep Disorders. In: Attarian H, Viola-Saltzman M, editors. Sleep Disorders in Women: A Guide to Practical Management. Cham: Springer International Publishing; 2020. p. 65-81.

17. Krishnan V, Collop NA. Gender differences in sleep disorders. Curr Opin Pulm Med. 2006;12(6):383-9.

18. Suh S, Cho N, Zhang J. Sex Differences in Insomnia: from Epidemiology and Etiology to Intervention. Curr Psychiatry Rep. 2018;20(9):69.

19. Zhang B, Wing Y-K. Sex Differences in Insomnia: A Meta-Analysis. Sleep. 2006;29(1):85-93.

20. Bublitz M, Adra N, Hijazi L, Shaib F, Attarian H, Bourjeily G. A Narrative Review of Sex and Gender Differences in Sleep Disordered Breathing: Gaps and Opportunities. Life. 2022;12(12):2003.

21. Meers J, Stout-Aguilar J, Nowakowski S. Chapter 3 - Sex differences in sleep health. In: Grandner MA, editor. Sleep and Health: Academic Press; 2019. p. 21-9.

22. Jonasdottir SS, Minor K, Lehmann S. Gender differences in nighttime sleep patterns and variability across the adult lifespan: a global-scale wearables study. Sleep. 2020;44(2).

23. Madrid-Valero JJ, Kirkpatrick RM, González-Javier F, Gregory AM, Ordoñana JR. Sex differences in sleep quality and psychological distress: Insights from a middle-aged twin sample from Spain. Journal of Sleep Research.n/a(n/a):e13714.

24. Mallampalli MP, Carter CL. Exploring sex and gender differences in sleep health: a Society for Women's Health Research Report. J Womens Health (Larchmt). 2014;23(7):553-62.

25. Mong JA, Baker FC, Mahoney MM, Paul KN, Schwartz MD, Semba K, et al. Sleep, rhythms, and the endocrine brain: influence of sex and gonadal hormones. J Neurosci. 2011;31(45):16107-16.

26. Mong JA, Cusmano DM. Sex differences in sleep: impact of biological sex and sex steroids. Philos Trans R Soc Lond B Biol Sci. 2016;371(1688):20150110.

27. Monterrosa-Castro Á, Portela-Buelvas K, Salguedo-Madrid M, Mo-Carrascal J, Leidy CD-M. Instruments to study sleep disorders in climacteric women. Sleep Science. 2016;9(3):169-78.

28. Baker FB. The basics of item response theory. Madison, WI, USA: ERIC Clearinghouse on Assessment and Evaluation; 2001.

29. De Boeck P, Wilson M. Explanatory item response models: A generalized linear and nonlinear approach: Springer; 2004.

30. Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. Journal of Educational measurement. 1990;27(4):361-70.

31. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. Sleep Med Rev. 2016;25:52-73.

32. Santos KD, Fischer FM, de Andrade DF, Barbetta PA, Alves Cardoso MR, de Gouveia Vilela RA. Shift work and sex are variants across Karolinska Sleep Questionnaire and the application of linking analysis as a step forward using Item Response Theory. Chronobiol Int. 2019;36(1):42-56.

33. van Kooten J, Terwee CB, Luijten MAJ, Steur LMH, Pillen S, Wolters NGJ, et al. Psychometric properties of the Patient-Reported Outcomes Measurement Information System (PROMIS) Sleep Disturbance and Sleep-Related Impairment item banks in adolescents. J Sleep Res. 2021;30(2):e13029.

34. Knott RJ, Lorgelly PK, Black N, Hollingsworth B. Differential item functioning in quality of life measurement: An analysis using anchoring vignettes. Soc Sci Med. 2017;190:247-55.

35. Saltychev M, Widbom-Kolhanen SS, Pernaa KI. Sex-related differential item functioning of neck disability index. Disabil Rehabil. 2023:1-7.

36. Katajapuu N, Laimi K, Heinonen A, Saltychev M. Floor and ceiling effects of the World Health Organization Disability Assessment Schedule 2.0 among patients with chronic musculoskeletal pain. Int J Rehabil Res. 2019;42(2):190-2.

37. Bourion-Bedes S, Schwan R, Laprevote V, Bedes A, Bonnet JL, Baumann C. Differential item functioning (DIF) of SF-12 and Q-LES-Q-SF items among french substance users. Health Qual Life Outcomes. 2015;13:172.

38. Cavanagh A, Wilson CJ, Caputi P, Kavanagh DJ. Symptom endorsement in men versus women with a diagnosis of depression: A differential item functioning approach. Int J Soc Psychiatry. 2016;62(6):549-59.

39. Verdam MGE, Oort FJ, Sprangers MAG. Item bias detection in the Hospital Anxiety and Depression Scale using structural equation modeling: comparison with other item bias detection methods. Qual Life Res. 2017;26(6):1439-50.

40. Brzoska P. Assessment of quality of life in individuals with chronic headache. Psychometric properties of the WHOQOL-BREF. BMC Neurol. 2020;20(1):267.

FIGURE LEGENDS

Figure 1. Test characteristic curve in both sexes together

Figure 2. Item information functions of JSS items grouped by sex

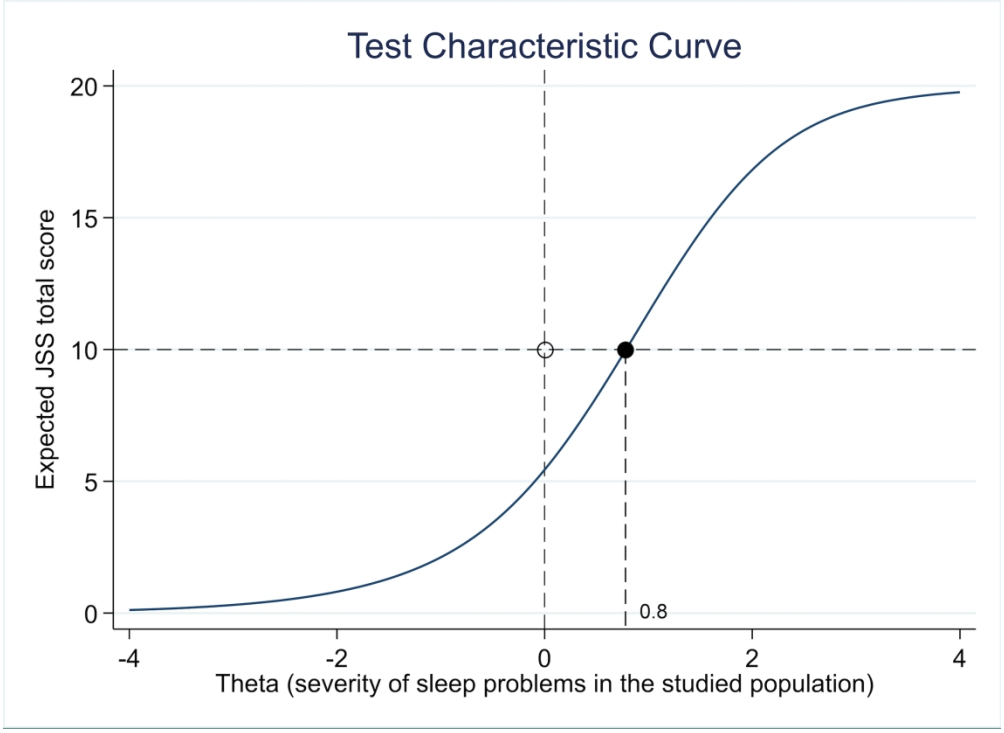

Figure 1. Test characteristic curve in both sexes together

87x63mm (762 x 762 DPI)

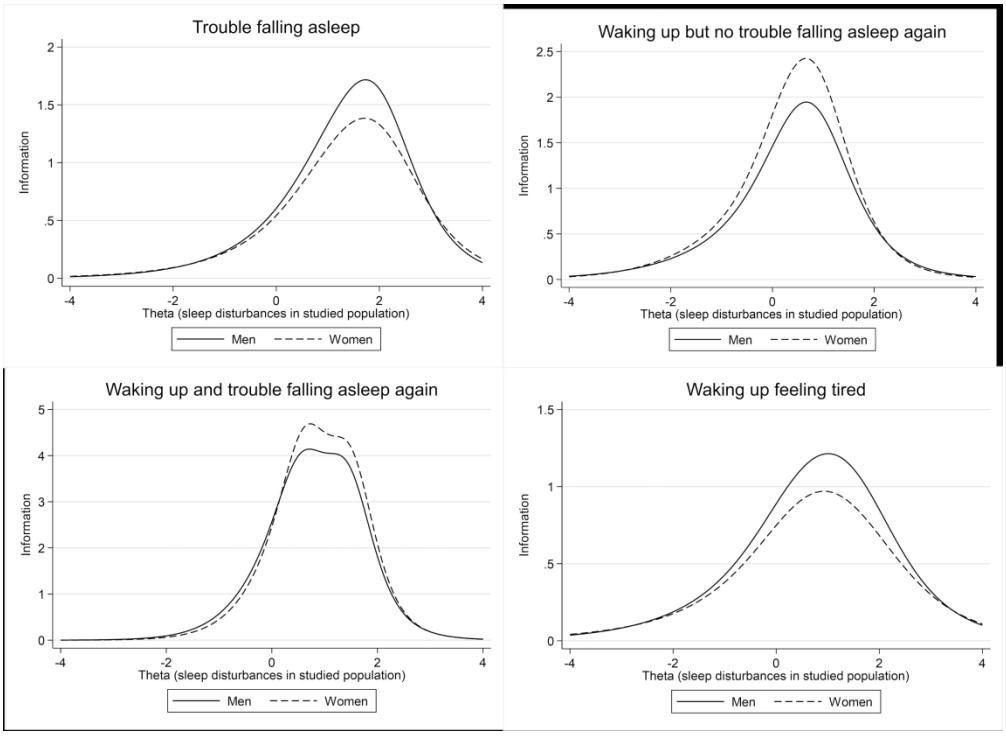

Figure 2. Item information functions of JSS items grouped by sex

147x107mm (762 x 762 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1-2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1-2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	3-4
Methods			
Study design	4	Present key elements of study design early in the paper	5-6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	5-6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5
Bias	9	Describe any efforts to address potential sources of bias	5-6
Study size	10	Explain how the study size was arrived at	5-6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5-6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5-6
		(*b*) Describe any methods used to examine subgroups and interactions	5-6
		(*c*) Explain how missing data were addressed	
		(*d*) If applicable, explain how loss to follow-up was addressed	5-6
		(*e*) Describe any sensitivity analyses	5-6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	5
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7
		(b) Indicate number of participants with missing data for each variable of interest	7
		(c) Summarise follow-up time (eg, average and total amount)	5
Outcome data	15*	Report numbers of outcome events or summary measures over time	7
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear	7

		which confounders were adjusted for and why they were included	
		(*b*) Report category boundaries when continuous variables were categorized	7
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	8
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	8-9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8-9
Generalisability	21	Discuss the generalisability (external validity) of the study results	9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	14

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.