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

Juhani Juhola [ID] ,[1] Jari P A Arokoski,[2,3] Jenni Ervasti [ID] ,[4] Mika Kivimäki,[5,6] Jussi Vahtera [ID] ,[7,8] Saana Myllyntausta,[9] Mikhail Saltychev [ID] [1]

For numbered affiliations see end of article.

Correspondence to
Juhani Juhola;
juhani.juhola@gmail.com

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 study (2015–2017).

Participants 77 967 employees in the Finnish public sector, with a mean age of 51.9 (SD 13.1) years and 82% women.

Outcome measures Item response theory estimates: difficulty and discrimination parameters of the JSS and differences in these parameters between men and women.

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 was moderate to high. For the JSS composite score, overall discrimination ability was moderate (0.98, 95% CI 0.97 to 0.99). Mild uniform DIF (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 of the respondents 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%, with no possibility of analysing 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 are a cost-efficient means to detect and grade sleep disturbances. The Jenkins Sleep Scale (JSS) was developed in 1998 as a brief and standardised questionnaire for sleep disturbances.[2] It is one of the most frequently 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 in patients with rheumatoid arthritis,[7] psoriatic arthritis,[6] ankylosing spondylitis,[5] fibromyalgia[10 11] and chest pain,[12] as well as in postcardiac 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 ranging from 0.7 to 0.9. Several studies have also assessed the internal consistency of the JSS in a general population, reporting respectively good to excellent Cronbach's alpha of 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 it 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 the item response theory or Rasch analysis. The 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 the 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, the 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 in 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 of developing 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 night-time impairment in women.[16] Sex differences in 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 night-time 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 to other causes.[16–18 20–26] Most of the previous studies on the topic have focused on sex-related and age-related differences in the prevalence and incidence of insomnia, while milder sleep disturbances have been less studied.

Previous studies have suggested that 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 night-time 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 organisations in 2015 (hospitals) and in 2016 (municipalities). There were 76 760 employees eligible for these surveys, of whom 53 505 (70%) responded. In addition, data were used from the 2017 survey sent to people who 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 whom 27 631 (57%) responded. There was 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 used in scientific research.

All 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 2. The level of physical activity was calculated from the survey responses and converted into a metabolic equivalent of task per hour per week (MET-hour/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 dichotomised 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, and waking up feeling tired (ie, waking up after the usual amount of sleep feeling tired and worn out). Each item was rated on a Likert-like scale from 0 to 5, where 0 is 'never', 1 is '1–3 days', 2 is 'about 1 night/week', 3 is '2–4 nights/week', 4 is '5–6 nights/week' and 5 is 'almost every night'. The total score is a simple sum of the scores of all four items and ranges from 0 ('no sleep problems') to 20 ('most sleep problems'). A score of ≤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 SD. The results were accompanied by 95% CI or two-tailed p values, when

appropriate. Using the item response theory, the average level of the 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 with 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. 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 2 or 3. 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 0.01–0.34 was considered 'none' (a completely level regression curve) or 'very low'; 0.35–0.64 was considered 'low'; 0.65–1.34 was considered 'moderate'; 1.35–1.69 was considered 'high'; and a discrimination of >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% CIs. The item characteristic curves for all four items are available from the corresponding author on request.

DIF is a statistical characteristic of a scale item (here counted for each of the four items included in the 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 non-uniform 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 non-uniform DIF is observed when the direction of difference between groups varies at different levels of sleep problems (eg, 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 V.17 statistical software.

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, BMI 26.2 (SD 4.7) kg/m², physical activity 29.6 (25.3) METs/week and alcohol

Table 1 Descriptive characteristics of the study sample

Variable	Men		Women		Entire sample	
	n	**%**	**n**	**%**	**n**	**%**
Total					77 967	100
Sex						
Male	100	0	0	100	14 349	18
Female	0	100	100	0	63 618	82
Sleep						
≤7 hours/night	10 779	75	45 235	71	56 014	72
>7 hours/night	3570	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/m²	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
Jenkins Sleep Scale, points	5.8	4.6	6.6	4.8	6.4	4.8

MET-hour/week, metabolic equivalent of task per hour per week.

Table 2 Difficulty coefficients of the JSS items in both sexes together (N=77967)

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

JSS, Jenkins Sleep Scale.

consumption 50.1 (SD 91.3) g/week (equivalent to around four units of alcohol per week). Of the respondents, 56 014 (72%) were sleeping 7 or less hours per night. The mean JSS total score was 6.4 (SD 4.8) points. Of the respondents, 12 629 (16%) had a JSS total score of more than 11.

Difficulty parameter of the 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 0 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 1 point. This shift towards underestimation was, however, mild. The same mild shift towards underestimation of the sleep problems was seen for both sexes (table 3 and figure 1).

Discrimination parameter of the 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).

DIF of the 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 the curves were close to uniform for all the items.

DISCUSSION

In this survey-based, cross-sectional study among 77967 employees in the Finnish public sector, there were minor differences in the psychometric properties of the 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 the items 'trouble falling asleep' and 'waking up feeling tired' and among women for the 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 differences in the incidence of these disorders between men and women. Women have more hormone-related sleep disorders[16–18 20–26] and also restless legs 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 pharmacological treatments may require different dosages based on sex.[16]

The generalisability of the results might be weakened by the sex imbalance of the studied sample (women 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,

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

JSS, Jenkins Sleep Scale.

the mean age of the 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 the JSS has never undergone a full linguistic validation process, which might affect its equivalency with an English version. The response rate in the surveys was 57% in 2015–2016 and 70% in 2017. No analyses were conducted on whether the demographic characteristics of the 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 the 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 is also the first study to explore the DIF of the JSS. However, the results of this study reflect previously observed differences in the amount and severity of sleep difficulties
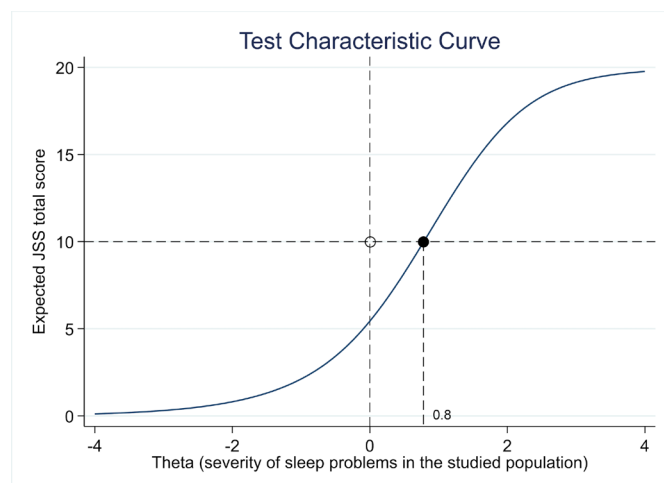
Test Characteristic Curve

Figure 1 Test characteristic curve in both sexes together. JSS, Jenkins Sleep Scale.

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 shown similar sex-related inconsistencies.[31] The DIF has been reported for the Karolinska Sleep Questionnaire.[32] Also, PROMIS (Patient-Reported Outcomes Measurement Information System), a very popular standard general PROM, has demonstrated an age-related DIF regarding sleep.[33]

The sex-related DIF of PROMs 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 the comparison should separately be performed by sex groups. This can be particularly true when, in addition to a composite score, the 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 (eg, sleep apnoea and disordered breathing or cardiovascular and metabolic disorders) may show results which are different from the present ones.
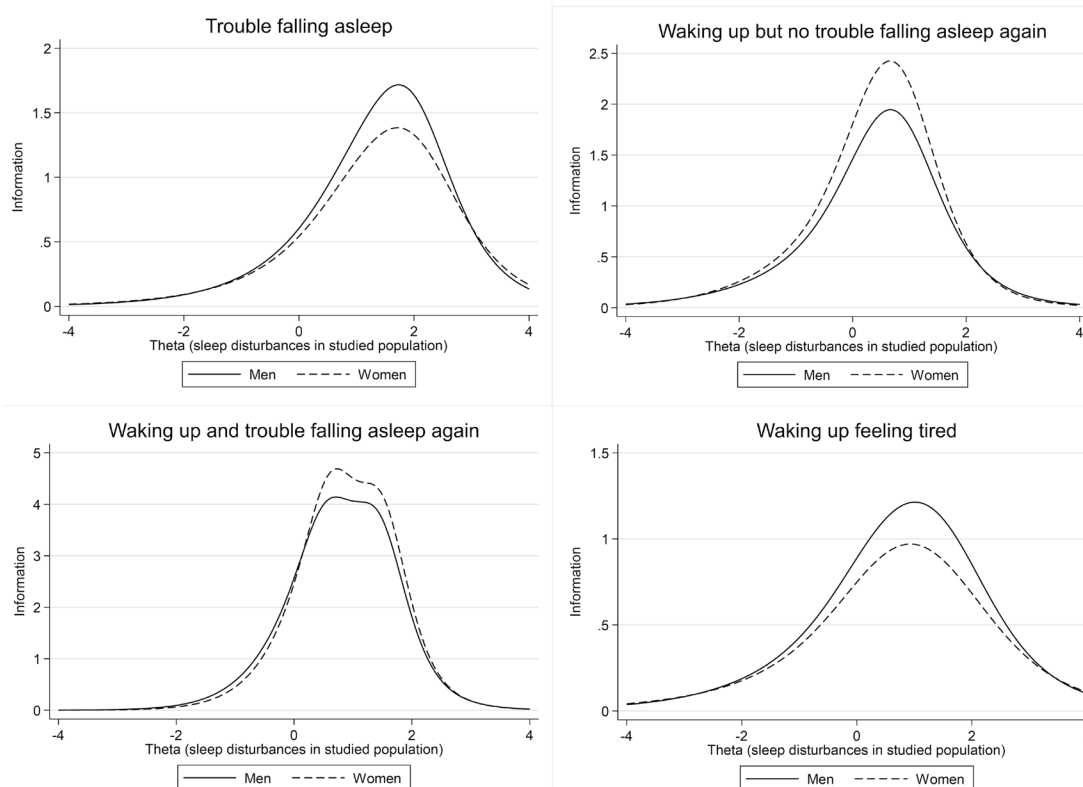

Figure 2 Item information functions of the JSS items grouped by sex. JSS, Jenkins Sleep Scale.

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 severities 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 are statistically significant, they are probably negligible when applied to clinical situations.

Author affiliations
¹Department of Physical and Rehabilitation Medicine, Turku University Hospital and University of Turku, Turku, Finland
²Department of Physical and Rehabilitation Medicine, Helsingin yliopistollinen Keskussairaala, Helsinki, Finland
³Helsingin yliopisto, Helsinki, Finland
⁴Finnish Institute of Occupational Health, Helsinki, Finland
⁵University of Helsinki Faculty of Medicine, Helsinki, Finland
⁶Department of Epidemiology and Public Health, University College London, London, UK
⁷Department of Public Health, Turun Yliopisto, Turku, Finland
⁸Finnish Institute of Occupational Health, Turku, Finland
⁹University of Turku, Turku, Finland

Twitter Jenni Ervasti @JenniErvasti1

Contributors All authors substantially contributed to the conception and design of the work and to drafting the work, revised it critically for important intellectual content, interpreted the data and finally approved the version published. All 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 the guarantor.

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 This study involves human participants and was approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa (registration number 60/13/03/00/2011 and 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. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

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.

ORCID iDs
Juhani Juhola http://orcid.org/0000-0002-7592-8183
Jenni Ervasti http://orcid.org/0000-0001-9113-2428
Jussi Vahtera http://orcid.org/0000-0002-6036-061X
Mikhail Saltychev http://orcid.org/0000-0003-1269-4743

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;21:474–80.
2. Jenkins CD, Stanton BA, Niemcryk SJ, *et al*. A scale for the estimation of sleep problems in clinical research. *J Clin Epidemiol* 1988;41:313–21.
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:S0022-3999(19)30714-7.
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;29:673–82.
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;22:275–9.
6. Duruöz MT, Erdem D, Gencer K, *et al*. Validity and reliability of the Turkish version of the Jenkins sleep scale in Psoriatic arthritis. *Rheumatol Int* 2018;38:261–5.
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;33:160–7.
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. *LP* 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. *JPPCM* 2018;4:04–6. 10.5530/jppcm.2018.1.2 Available: http://jppcm.org/v4/i1
10. Unal-Ulutatar C, Ozsoy-Unubol T. Psychometric properties of Turkish version of Jenkins sleep scale in Fibromyalgia syndrome. *Adv Rheumatol* 2020;60:22.
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: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 Nurs* 2006;5:7:1–10.:.
13. Nasermoaddeli A, Sekine M, Kumari M, *et al*. Association of sleep quality and free time leisure activities in Japanese and British civil servants. *J Occup Health* 2005;47:384–90.
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;77:344–50.
15. Carrier J, Semba K, Deurveilher S, *et al*. Sex differences in age-related changes in the sleep-wake cycle. *Front Neuroendocrinol* 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;12:383–9.
18. Suh S, Cho N, Zhang J. Sex differences in insomnia: from epidemiology and etiology to intervention. *Curr Psychiatry Rep* 2018;20:69.
19. Zhang B, Wing Y-K. Sex differences in insomnia: A meta-analysis. *Sleep* 2006;29:85–93.
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 (Basel)* 2022;12:2003:12.:.
21. Meers J, Stout-Aguilar J, Nowakowski S. Chapter 3 - sex differences in sleep health. In: Grandner MA, ed. *Sleep health*. Academic Press, 2019: 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* 2021;44.
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. *J Sleep Res* 2023;32:e13714. 10.1111/jsr.13714 Available: https://onlinelibrary.wiley.com/toc/13652869/32/2
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;23:553–62.
25. Mong JA, Baker FC, Mahoney MM, *et al*. Sleep, rhythms, and the endocrine brain: influence of sex and Gonadal hormones. *J Neurosci* 2011;31:16107–16.

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;371:20150110.

27 Monterrosa-Castro Á, Portela-Buelvas K, Salguedo-Madrid M, *et al*. Instruments to study sleep disorders in Climacteric women. *Sleep Sci* 2016;9:169–78.

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

29 DeP, 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. *J Educational Measurement* 1990;27:361–70. 10.1111/j.1745-3984.1990.tb00754.x Available: https://onlinelibrary.wiley.com/toc/17453984/27/4

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.

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;36:42–56.

33 van Kooten JAMC, 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:e13029.

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.

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

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;42:190–2.

37 Bourion-Bédès 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.

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;62:549–59.

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;26:1439–50.

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