PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Internal consistency and factor structure of Jenkins Sleep Scale – cross-sectional cohort study amongst 80,000 adults
AUTHORS	Juhola, Juhani; Arokoski, J. P. A.; Ervasti, Jenni; Kivimäki, Mika; Vahtera, Jussi; Myllyntausta, Saana; Saltychev, M

VERSION 1 – REVIEW

REVIEWER	Kariem Elhadd and Andrew J Larner Walton Centre for Neurology and Neurosurgery, Liverpool, UK
REVIEW RETURNED	11-Sep-2020

GENERAL COMMENTS	Whilst a Finnish language translation of a sleep assessment instrument may find limited use globally, nevertheless the large cohort examined in this study prompts confidence in the finding of JSS as an internally consistent and unidimensional scale. Hence it may be deemed suitable for use as a screening instrument for sleep disturbance in population based studies. However, little is said about the limitations of JSS. Like any brief screening instrument, it has shortcomings, specifically the inability to address the spectrum of sleep disorders (see Shahid A, Wilkinson K, Marcu S, Shapiro CM. Jenkins Sleep Scale. In: Shahid A, Wilkinson K, Marcu S, Shapiro CM (eds). STOP, THAT and one hundred other sleep scales. New York: Springer, 2011:203-4). Hence it can only be used as a preliminary screener of sleep disturbance. Why is there such a gender imbalance (F:M = >4:1) in the study cohort? As this is a potential limitation to the generalisability of the findings, this necessitates some explanation. It is surely unlikely to be related to the overall population, so is it related to response rate (F>M), or are fewer men working as employees in the population surveyed? Likewise, why is the mean age of study participants (52 years) in the last third of the working life span? Does this reflect the demographic of the working population in Finland? Please expand on this if possible. Is the Finnish version of JSS already published, or is this a new translation? If the latter, surely methods of translation and back translation need to be described? This is particularly important if “some differences might have occurred due to the linguistic variability between the two translations” (P8 of 19, line 46). Other points:

	P3 of 19, line 27: “(95% CI 0.xx-0.xx).”. Insert the confidence intervals. P4 of 19, line 5: “in 1998” to read “in 1988”. The index publication by Jenkins et al. was in 1988! P4 of 19, lines 10-12: this listing of applications of JSS might also include patients with cognitive disorders (e.g. Neurodegener Dis Manag 2018;8:377-83 and J Sleep Disord Ther 2018;7:291) and with epilepsy (e.g. J Sleep Disord Ther 2019;8:1). P5 of 19, lines 36-38: “The score of 11 is a cut-off – a score <12 is defined as little of sleep disturbances and a score >11 is understood as high frequency of sleep disturbances[16].”. There are other ways to dichotomise the scale, see ref. 1 (yes to any one of the questions for more than 15 nights in the previous 4 weeks). P5 of 19, lines 56-58: a reference to the classification of Cronbach alpha values would be helpful. P6 of 19, lines 17-18: a reference to the classification of correlation coefficient values would be helpful. P7 of 19, line 4: “Of 81,136 respondents”. If this represents around 70% response rate (as at P5 of 19, line 8), it would be helpful to know if there are any data on the demographic factors of the responders versus the non-responders to see if there might be any bias. P10 of 19, line 6: “Jenkin’s” to read “Jenkins”. P14 of 19, line 23: no journal name, volume, or pagination. Table 3: some explanation of these numbers would be helpful, e.g. is BIC value of 1,037,000 a good, bad, or indifferent as an information criterion? Do these values have classification ranges? Figure 2: the numbers in the bottom half of the figure (0.55, 0.56, 0.39, 0.55) need some explanation (I don’t see them in any of the tables), as does ε. K Elhadd AJ Larner Walton Centre for Neurology and Neurosurgery, Liverpool, UK

REVIEWER	Markus Jansson-Fröjmark Karolinska Institutet, Sweden
REVIEW RETURNED	22-Oct-2020

GENERAL COMMENTS	There are several strengths to this paper, particularly the large sample and a concise text. The following comments need to be addressed in a revision: • Abstract: (1) It appears as if the confidence intervals are missing alpha for the JSS. (2) Since the JSS is not a diagnostic instrument, the term “sleep disorders” in the conclusions (and under

	"Strengths and limitations of this study" and on other instances in the paper) part is a bit strong. I would suggest the authors rephrase to "sleep problems" or similar. • The introduction is short and concise. The need to explore the JSS in large samples is underscored as well as to do so in Finland. First, I am not certain what the rationale is to examine the JSS in Finnish; e.g., would the authors expect different findings to emerge, and, if so, why? Second, I wonder whether previous research on the JSS also has methodological limitations that might warrant a new study. This point is lacking in the introduction. • Methods: The population is only briefly described and the sampling procedure not all. I recommend the authors to add more information on these two aspects of the study as well as on the FPS sub-cohorts. The MET methodology was unfamiliar to me, and probably also to a segment of future readers; I would advise the readers to add more information about the MET. • Statistical analysis: In the EFA, the authors use eigenvalues to determine the number of factors. Several methodological papers have shown that there are stronger options, particularly the use of parallel analysis. Therefore, I ask the authors to reconsider using the parallel analysis instead as the primary method. • Results: I note that the authors did not use measurement variance estimations. I believe that such analyses, e.g., comparing women/men or shorter/longer sleep time, would be informative. This comment is also related to one of the limitations that the authors raise, namely that the generalizability might be compromised. To establish measurement invariance would be one way to reduce the issue of potential lack of generalizability. • Discussion: First, the authors mention the need to carry out IRT analyses. Why was that not executed in the current paper? This would have enabled a broader perspective on how the JSS operates. Second, the authors argue that the JSS may be used as a measure in the healthy, working-age population. I recommend the authors to be a bit more specific concerning this notion, e.g., for whom and under what circumstances. Relative to other similar measures (e.g., the Insomnia Severity Index), what are the advantages and disadvantages of using the JSS?

VERSION 1 – AUTHOR RESPONSE

Response to the comments made by the Reviewer 1

Comment 1 However, little is said about the limitations of JSS. Like any brief screening instrument, it has shortcomings, specifically the inability to address the spectrum of sleep disorders (see Shahid A, Wilkinson K, Marcu S, Shapiro CM. Jenkins Sleep Scale. In: Shahid A, Wilkinson K, Marcu S, Shapiro CM (eds). STOP, THAT and one hundred other sleep scales. New York: Springer, 2011:203-4). Hence it can only be used as a preliminary screener of sleep disturbance.
Response 1

We have now added the following text to the Introduction section along with the suggested reference:

“Like any brief screening instrument, the JSS has shortcomings, specifically the inability to address the spectrum of sleep disorders. Hence it can only be used as a preliminary screener of sleep disturbance.”

Comment 2

Why is there such a gender imbalance (F:M = >4:1) in the study cohort? As this is a potential limitation to the generalisability of the findings, this necessitates some explanation. It is surely unlikely to be related to the overall population, so is it related to response rate (F>M), or are fewer men working as employees in the population surveyed?

Response 2

We have now added the following text to the “strengths and weaknesses” paragraph of the Discussion section:

“The generalizability of the results might be weakened by the gender disbalance of the studied cohort (women were predominated). This disbalance was due to the fact that fewer men are involved in the studied areas of public sector”.

Comment 3

Likewise, why is the mean age of study participants (52 years) in the last third of the working life span? Does this reflect the demographic of the working population in Finland? Please expand on this if possible.

Response 3

We have now added the following text to the “strengths and weaknesses” paragraph of the Discussion section:

“Also, the mean age of study participants was 52 years and, therefore, the results described, in the first instance, people in the last third of their working life span.”

Comment 4

Is the Finnish version of JSS already published, or is this a new translation? If the latter, surely methods of translation and back translation need to be described? This is particularly important if

“some differences might have occurred due to the linguistic variability between the two translations” (P8 of 19, line 46).
Response 4 We have now added the following text to the “strengths and weaknesses” paragraph of the Discussion section: “While 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”.

Comment 5 P3 of 19, line 27: “(95% CI 0.xx-0.xx).”. Insert the confidence intervals.
Response 5 The typo has now been corrected.

Comment 6 P4 of 19, line 5: “in 1998” to read “in 1988”. The index publication by Jenkins et al. was in 1988!
Response 6 The typo has now been corrected.

Comment 7 P4 of 19, lines 10-12: this listing of applications of JSS might also include patients with cognitive disorders (e.g. Neurodegener Dis Manag 2018;8:377-83 and J Sleep Disord Ther 2018;7:291) and with epilepsy (e.g. J Sleep Disord Ther 2019;8:1).
Response 7 The text of the Introduction section has now been changed as follows. Also, new references have now been added.

"The JSS has been translated in several languages and found to be valid and reliable amongst patients with different health problems including rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, fibromyalgia, chest pain, post cardiac surgery patients, patients with cognitive disorders and epilepsy."

Comment 8

P5 of 19, lines 36-38: "The score of 11 is a cut-off – a score <12 is defined as little

of sleep disturbances and a score >11 is understood as high frequency of sleep disturbances[16].". There are other ways to dichotomise the scale, see ref. 1 (yes to any one of the questions for more than 15 nights in the previous 4 weeks).

Response 8

We have now modified the text of the Methods section as follows:

"Another way to dichotomize the JSS is considering sleep difficulties being present if there is at least one "yes" response (>15 nights in the previous 4 weeks) to any item."

Comment 9

P5 of 19, lines 56-58: a reference to the classification of Cronbach alpha values would be helpful.

Response 9

The reference has now been added as suggested:

"George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: Allyn & Bacon."

Comment 10

P6 of 19, lines 17-18: a reference to the classification of correlation coefficient values would be helpful.

Response 10

The reference has now been added as suggested:

Haldun Akoglu. User's guide to correlation coefficients. Turk J Emerg Med. 2018 Sep; 18(3): 91–93.

Comment 11

P7 of 19, line 4: “Of 81,136 respondents”. If this represents around 70% response rate (as at P5 of 19, line 8), it would be helpful to know if there are any data on the demographic factors of the responders versus the non-responders to see if there might be any bias.

Response 11

We have now added the following text to the “strengths and weaknesses” paragraph of the Discussion section:

“The response rate was 70% and there was no analysis if the non-respondents’ demographic characteristics might affect the results.”

Comment 12

P10 of 19, line 6: “Jenkin’s” to read “Jenkins”.

Response 12

The typo has now been corrected.

Comment 13

P14 of 19, line 23: no journal name, volume, or pagination.

Response 13

The typo has now been corrected.

Comment 14

Table 3: some explanation of these numbers would be helpful, e.g. is BIC value of 1,037,000 a good, bad, or indifferent as an information criterion? Do these values have classification ranges?

Response 14

The statistical analysis section contains the sentence:

"The AIC and BIC were considered good if they were close to 1.0."

Comment 15

Figure 2: the numbers in the bottom half of the figure (0.55, 0.56, 0.39, 0.55) need some explanation (I don't see them in any of the tables), as does ε.

Response 15

The explanations have now been added to the footnote of Fig 2 as follows:

"ε-circles represent a measurement error associated with an observed variable (variance that is predicted by the latent factor); estimates placed between ε-errors and observed variables represent the amount of variance in higher level data that can be explained by a particular variable"

Response to the comments made by the Reviewer 2

Comment 1

Abstract: (1) It appears as if the confidence intervals are missing alpha for the JSS. (2) Since the JSS is not a diagnostic instrument, the term "sleep disorders" in the conclusions (and under "Strengths and limitations of this study" and on other instances in the paper) part is a bit strong. I would suggest the authors rephrase to "sleep problems" or similar.

Response 1

The typo has now been corrected. The term "sleep disorders" has now been changed to "sleep difficulties" through the entire text.

Comment 2

The introduction is short and concise. The need to explore the JSS in large samples is underscored as well as to do so in Finland. First, I am not certain what the rationale is to examine the JSS in Finnish; e.g., would the authors expect different findings to emerge, and, if so, why? Second, I

wonder whether previous research on the JSS also has methodological limitations that might warrant a new study. This point is lacking in the introduction.

Response 2

We have now expended the following text in the Introduction section:

"Overall, there is uncertainty concerning the psychometric behavior of the JSS especially regarding its factor structure in healthy and/or general populations. Concerning a general population, previous research mostly focused on the internal consistency of JSS and its reliability. Instead, other important points, like e.g. factors structure, remained practically unknown. Additionally, the psychometric properties of Finnish translation of the JSS have not been studied yet."

Comment 3

Methods: The population is only briefly described and the sampling procedure not all. I recommend the authors to add more information on these two aspects of the study as well as on the FPS sub-cohorts. The MET methodology was unfamiliar to me, and probably also to a segment of future readers; I would advise the readers to add more information about the MET.

Response 3

We have now modified/added the following text to the Methods section:

"The data were derived from the Finnish Public Sector (FPS) study, an on-going prospective cohort study of employees in the municipal services of 10 Finnish towns and 21 public hospitals. The eligible population from the register cohort of FPS (n=151 618) included those who had been employed for a minimum of 6 months at the participating organisations between 1991 and 2005. Employers' records have been used to identify the eligible employees for a nested survey cohort to whom questionnaire surveys have been repeated every 4 years since 2000 [15]. For this study, the data were sourced from the survey in 2016 – 2017 administered to the FPS sub-cohorts (average response rate 70%). 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.

Also, we have now added a new reference:

Márcio de Almeida Mendes. Metabolic equivalent of task (METs) thresholds as an indicator of physical activity intensity. PLoS One. 2018; 13(7): e0200701. PMID: 30024953

Comment 4

Statistical analysis: In the EFA, the authors use eigenvalues to determine the number of factors. Several methodological papers have shown that there are stronger options, particularly the use of parallel analysis. Therefore, I ask the authors to reconsider using the parallel analysis instead as the primary method.

Response 4

We have employed a parallel analysis as stated in the Statistical analysis section and in the Fig.1. We apologize for not reporting on that in the Results. It has now been added to the Results section as follows:

“The parallel analysis of scree plot confirmed the unidimensional structure of JSS.”

Comment 5

Results: I note that the authors did not use measurement variance estimations. I believe that such analyses, e.g., comparing women/men or shorter/longer sleep time, would be informative. This comment is also related to one of the limitations that the authors raise, namely that the generalizability might be compromised. To establish measurement invariance would be one way to reduce the issue of potential lack of generalizability.

Response 5

We apologize for not completely understanding the question. If the question is about gender- or sleep time- differential item functioning then it was out of the scope of the study.

Comment 6

Discussion: First, the authors mention the need to carry out IRT analyses. Why was that not executed in the current paper? This would have enabled a broader perspective on how the JSS operates.

Response 6

The IRT analysis was outside the scope of this study. We are considering such an analysis and we feel that such a complex angle deserves a full-scale paper for its own.

Comment 6

Discussion: Second, the authors argue that the JSS may be used as a measure in the healthy, working-age population. I recommend the authors to be a bit more specific concerning this notion, e.g., for whom and under what circumstances. Relative to other similar measures (e.g., the Insomnia Severity Index), what are the advantages and disadvantages of using the JSS?
Response 6 We have now added the following text to the Introduction section: “Comparing to other similar measures, the JSS is a short questionnaire focusing on roughly recognizing sleep difficulties. That is unlike to more complex scales, like Insomnia Severity Index, which quantify also the impact of sleep disturbance on the level of daily functioning.”. We have now modified the following text of the Conclusions section: “The JSS was found to be a unidimensional scale with good internal consistency. As such, the JSS may be recommended as an easy-to-do questionnaire instrument for the screening for sleep difficulties in a healthy working-age population.”

VERSION 2 – REVIEW

REVIEWER	Markus Jansson-Fröjmark Karolinska Institutet, Sweden
REVIEW RETURNED	26-Nov-2020

GENERAL COMMENTS	The authors have definitely made reasonable revisions since the first version and improved the paper in general. There are just a few minor aspects of the paper that need further revision: 1. Relating to reviewer 1's comment 9: the reference chosen for the classification of Cronbach alpha values appears to be a textbook. My recommendation is that the authors add a primary source instead. 2. Relating to reviewer 1's comment 11: it is now stated that no attrition analysis was carried out. If there are any data related to the non-responders, I would advise the authors to investigate potential differences between responders and non-responders. 3. Relating to my comment 4: the parallel analysis and scree plot are two separate methods. So I am confused as to why it now reads: "The parallel analysis of scree plot...". 3. Relating to my comment 5: I apologize for being too vague in my previous feedback. I did not mean to say that I advised the authors to investigate differential item functioning, although it is an interesting option in itself. What I meant to say was that exploring measurement invariance in relation to the CFA might provide enhanced knowledge of whether the JSS is performing in a similar way across groups. So using a group variable, such as gender or sleep time, would work. This reference describes the methodology further: Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis

	framework. Journal of Psychoeducational Assessment, 29(4), 347-363.

VERSION 2 – AUTHOR RESPONSE

Response to the comments made by the Reviewer 2

Comment 1 Relating to reviewer 1's comment 9: the reference chosen for the classification of Cronbach alpha values appears to be a textbook. My recommendation is that the authors add a primary source instead.
Response 1 We have now added a new refence: Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Research in Science Education 2018;48:1273–96.

Comment 2 Relating to reviewer 1's comment 11: it is now stated that no attrition analysis was carried out. If there are any data related to the non-responders, I would advise the authors to investigate potential differences between responders and non-responders.
Response 2 Unfortunately, the data on the nonrespondents are not available for the analysis. We have stated that in the "strengths and weaknesses" paragraph: "The response rate was 70% and there was no analysis of whether the non-respondents' demographic characteristics might affect the results."

Comment 3 Relating to my comment 4: the parallel analysis and scree plot are two separate methods. So I am confused as to why it now reads: "The parallel analysis of scree plot...".
Response 3 We have now expanded the report regarding the parallel analysis adding a new table (Table 2).

Comment 4 Relating to my comment 5: I apologize for being too vague in my previous feedback. I did not mean to say that I advised the authors to investigate differential item functioning, although it is an interesting option in itself. What I meant to say was that exploring measurement invariance in relation to the CFA might provide enhanced knowledge of whether the JSS is performing in a similar way across groups. So using a group variable, such as gender or sleep time, would work. This reference describes the methodology further: Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. Journal of Psychoeducational Assessment, 29(4), 347-363.
Response 4 We have considered the suggestion. Though an additional analysis may be of value, we would like to limit this paper to a conventional CFA analysis.

VERSION 3 – REVIEW

REVIEWER	Markus Jansson-Fröjmark Karolinska Institutet, Sweden
REVIEW RETURNED	17-Dec-2020

GENERAL COMMENTS	It is my opinion that the current resubmission is to be accepted for publication.