



# Validation of the Arabic version of the Epworth Sleepiness Scale

Anwar E. Ahmed <sup>a</sup>, Abdulhamid Fatani <sup>b</sup>, Abdullah Al-Harbi <sup>c</sup>,  
Abdullah Al-Shimemeri <sup>c</sup>, Yosra Z. Ali <sup>a</sup>, Salim Baharoon <sup>c</sup>,  
Hamdan Al-Jahdali <sup>c,\*</sup>

<sup>a</sup> Department of Epidemiology and Biostatistics, College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia

<sup>b</sup> College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia

<sup>c</sup> Department of Medicine, Pulmonary Division, Sleep Disorders Center, King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia

Received 7 January 2014; received in revised form 19 April 2014; accepted 23 April 2014  
Available online 2 July 2014

## KEYWORDS

Epworth Sleepiness Scale;  
Daytime sleepiness;  
Sleep disorder

**Abstract Background:** The Epworth Sleepiness Scale (ESS) is a questionnaire widely used in developed countries to measure daytime sleepiness and diagnose sleep disorders.

**Objective:** This study aimed to develop an ESS questionnaire for the Arabic population (ArESS), to determine ArESS internal consistency, and to measure ArESS test–retest reproducibility. It also investigated whether the normal range of ESS scores of healthy people in different cultures are similar.

**Methods:** The original ESS questionnaire was translated from English to Arabic and back-translated to English. In both the English and Arabic translations of the survey, ESS consists of eight different situations. The subject was asked to rate the chance of dozing in each situation on a scale of 0–3 with total scores ranging between 0 (normal sleep) and 24 (very sleepy). An Arabic translation of the ESS questionnaire was administered to 90 healthy subjects.

**Results:** Item analysis revealed high internal consistency within ArESS questionnaire (Cronbach's alpha = 0.86 in the initial test, and 0.89 in the retest). The test–retest intra-class correlation coefficient (ICC) shows that the test–retest reliability was substantially high: ICC = 0.86 (95% confidence interval: 0.789–0.909,  $p$ -value < 0.001). The difference in ArESS scores between the initial test and retest was not significantly different from zero (average difference =  $-0.19$ ,  $t = -0.51$ ,  $df = 89$ ,  $p$ -value = 0.611). In this study, the averages of the ESS scores

\* Corresponding author. Address: King Saud University for Health Sciences, Pulmonary Division, Sleep Disorders Center, King Abdulaziz Medical City, Riyadh, Saudi Arabia. Tel.: +966 12520088x17597, 17531 (work), mobile: +966 505224271.  
E-mail address: [Jahdali@gmail.com](mailto:Jahdali@gmail.com) (H. Al-Jahdali).

( $6.3 \pm 4.7$ , range 0–20 in the initial test and  $6.5 \pm 5.3$ , range 0–20 in the retest) are considered high in Western cultures.

*Conclusions:* The study shows that the ArESS is a valid and reliable tool that can be used in Arabic-speaking populations to measure daytime sleepiness. The current study has shown that the average ESS score of healthy Arabian subjects is significantly higher than in Western cultures.

© 2014 Ministry of Health, Saudi Arabia. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The daytime sleepiness problem is a common symptom of many sleep disorders, including obstructive sleep apnea. The Epworth Sleepiness Scale (ESS) questionnaire was developed by Johns as a simple, self-administered questionnaire for the assessment of daytime sleepiness [1–3]. Since its development in 1991, it has been used widely in clinical practice, sleep laboratory questionnaires, and research protocols as a simple rapid assessment of subjective sleepiness [2]. ESS is consistent with eight different situations, and the subject is asked to rate the probability of dozing in each situation on a scale of 0–3 (0 = no chance of dozing, 1 = slight chance of dozing, 2 = moderate chance of dozing, 3 = high chance of dozing; minimum score = 0, maximum score = 24), with total scores ranging between 0 (normal sleep) and 24 (very sleepy) [2].

The ESS questionnaire is developed in the English language and used among Western individuals; therefore, a direct translation may have some limitations for use in other countries, due to cultural and economic differences. Several studies have been conducted to translate and validate the ESS questionnaire into different languages. The ESS questionnaire has been validated and used by many non-English language speakers' countries, including: Spanish [4], German [5], Chinese [6,7], Japanese [8], Turkish [9], Italian [10], and Greek [11]. In healthy people who do not have evidence of which is considered a sleep disorder, the normal range of ESS scores as defined by the 2.5 and 97.5 percentiles was 0–10 with an average of  $4.6 \pm 2.8$  [12]. There are wide variations across different cultures; the prevalence of excessive daytime sleepiness (ESS > 10) in Saudi healthcare workers is 39.3% (Wali et al. [18]), which is considered high in Western cultures. For instance, the prevalence of excessive daytime sleepiness in a normal Australian was found to be 11% at the cut-off point ESS > 10 [12]. It is not yet clear whether the normal range of ESS scores of healthy subjects in different cultures are similar.

It was found that there is a high level of internal consistency between the eight items in the ESS as measured by Cronbach's alpha, ranging from 0.74 to 0.88. Only one Arabic version of the ESS has been developed in three Lebanese sleep centers [16]. The investigators compared the ESS Arabic version between patients with sleep-related breathing diseases and healthy people. Their study revealed a high level of internal consistency; Cronbach's alpha was 0.76 and the intra-class correlation coefficient was 0.85. These investigators evaluated the internal consistency of the ESS scores at one time point [16]. The present study addresses the level of reliability of the ESS Arabic version at different settings, such as test and retest, where the same subjects are used to complete the survey on two different occasions. Further, the present study considers cultural differences in translating the current survey. The purpose of the study is to evaluate the reliability of the ArESS scores that were measured at different times by administering the same survey to healthy people. It was also investigated whether the normal range of ESS scores of healthy people in different cultures is similar.

## 2. Methodology

A test–retest design was conducted at King Abdulaziz Medical City–King Fahad National Guard Hospital (KAMC–KFNGH) in Riyadh, Saudi Arabia. The study was conducted between January and April 2013. This study was designed to evaluate and validate the ESS Arabic (ArESS) version as it measures daytime sleepiness. The original ESS questionnaire was translated from English to Arabic and back-translated to English by a professional translation office and then re-tested by two physicians and one sleep technologist, both of whom are fluent in Arabic and English. The ESS Arabic version and back-translation were compared by the two physicians to check for coherence and precision until both versions were considered completely interchangeable: conceptually and linguistically. Some cultural modifications must be considered in translating and evaluating the ArESS. One such modifica-

tion is seen in question 7: "Sitting quietly after lunch without alcohol." "Without alcohol" was deleted as alcohol consumption after lunch is a common habit among individuals in Western cultures, but even if used in the Arabic culture, it would be among a limited number of individuals and an exception to the cultural habit rather than the rule. In most of the Arabic society, alcohol consumption is considered religiously and socially unacceptable, and very few people could openly answer questions related to alcohol consumption. Other cultural considerations may be made in analyzing the responses of individuals, including gender-based prohibitions on operating motor vehicles and the relative importance of daily meals. For example, in Saudi Arabia, where women are prohibited from operating motor vehicles, female subjects are likely to rate higher in sleepiness in car- and traffic-related situations (like item 8 of the ESS) than men. Aiming to examine the internal consistency reliability of the Arabic version, the final version of ArESS (Appendix 1) was administered twice at an interval of four weeks to 90 healthy subjects who were willing to answer the questionnaires at four-week intervals. This group of subjects was comprised of students, health professionals, and other volunteers.

### 3. Statistical analysis

Descriptive statistics such as means and standard deviation (mean  $\pm$  SD) were used to describe the quantitative variables. On the other hand, a description of categorical variables was carried out by calculating frequencies and percentages,  $n$  (%). Cronbach's alpha coefficient [13] was used to determine the internal consistency of ArESS subscales. Scale items were considered to be homogeneous if Cronbach's alpha was above 0.70, but not higher than 0.90. Intra-class correlation coefficient [14] was used to address the question of whether the ArESS questionnaire on the first occasion (initial test) will show preserved subject differentiability on the second occasion (retest). A value of 0.81 or higher indicates perfect agreement between test and retest. The normal range of ArESS scores was defined by the 2.5 and 97.5 percentiles. All analyses were conducted using the Statistical Package for Social Sciences (SPSS Inc, Chicago, IL, USA), version 20.

### 4. Results

Ninety ( $N = 90$ ) participants were enrolled in the study. As shown in Table 1, the average age was

$29.8 \pm 12.6$  (SD) years, the average body mass index (BMI) was  $25.1 \pm 4.4$  (SD), and the average neck circumference was  $34.3 \pm 4.6$  (SD). About one third of participants (34.4%) were females. The majority of the participants (56.7%) were coffee drinkers, and 45.6% were tea drinkers. Of the 90 participants, 2.2% had no education, 5.6% had intermediate education, 21.1% had graduated from high school, and the majority of 71.1% had a university education. The majority of the participants (68.9%) were single, 26.7% were married, and 4.4% were divorced or widowed. Approximately 57% were university students, 29% were employed, 9% were housewives, and 5% were retired or unemployed. A majority (78.9%) of the respondents were nonsmokers, 6.7% were ex-smokers, while 14.4% were current smokers. The eight ArESS items were thoroughly reported by respondents and there were no missing item scores. The prevalence of excessive daytime sleepiness in healthy Saudi subjects was found to be 19% at the cut-off point  $>10$  in the initial test and 20% at the cut-off point  $>10$  in the retest.

In the initial test, the mean ArESS total score was a low  $6.3 \pm 4.7$  (SD) with a range of 0–20. The normal range of ArESS is 2.5 percentile = 0 and 97.5 percentile = 19, which is considered relatively high daytime sleepiness in Western cultures. The internal consistency reliability of the eight-item ArESS questionnaires was examined using Cronbach's alpha for test and retest, separately (Table 2). In the initial test, Cronbach's alpha indicates that the Arabic translation of the subscales (eight ArESS items) was consistent and homogeneous and that Cronbach's alpha was 0.86. Since the ArESS questionnaire shows good reliability on each occasion, the individual items within the ArESS questionnaire might not require modification.

In the retest, the mean ESS total score was  $6.5 \pm 5.3$  (SD) with a range of 0–22, and again it was found that the ArESS questionnaire is consistent and reliable (Cronbach's alpha was 0.89). (Table 2). The test–retest intra-class correlation coefficient (ICC) of absolute agreement revealed that test–retest reproducibility was high, and the ICC = 0.86 (95% confidence interval: 0.789–0.909,  $p$ -value  $< 0.001$ ). Additionally, the agreement between the test–retest can be obtained by testing whether the change in ArESS scores of test–retest was different from 0 (Table 3). The paired  $t$  test did not show a significant difference in ArESS scores for the initial test  $6.3 \pm 4.7$  (SD) and the retest  $6.5 \pm 5.3$  (SD) (average difference =  $-0.19$ ,  $t = -0.51$ ,  $df = 89$ ,  $p$ -value = 0.611). This indicates that the Arabic version of the ESS is a consistent and reliable measure to diagnose excessive day-

**Table 1** Characteristics from individuals completing reliability surveys ( $N = 90$ ).

Characteristics	Level	<i>n</i>	%
Gender	Female	31	34.4
Education	No education	2	2.2
	Intermediate	5	5.6
	High school	19	21.1
	University	64	71.1
Marital status	Single	62	68.9
	Married	24	26.7
	Widow	2	2.2
	Divorced	2	2.2
Employment	Employed	26	28.9
	Retired	3	3.3
	Unemployed	2	2.2
	Student	51	56.7
	Housewife	8	8.9
Smoke status	Non-smoker	71	78.9
	Ex-smoker	6	6.7
	Current smoker	13	14.4

time sleepiness. It was found that women do not have significantly higher sleepiness than men; the mean Arabic version of the ESS is the same across men and women ( $p$ -value = 0.394 for test and  $p$ -value = 0.861 for retest).

## 5. Discussion

The purpose of the study was to check the test and retest reproducibility of the Arabic ESS version and the internal consistency of the scale. Findings of the present investigation suggest that the Arabic version of the ESS has adequate internal consistency reliability in assessing sleep propensity: the internal consistency exceeded 0.70 on two occasions, initial test (Cronbach's alpha = 0.86) and retest (Cronbach's alpha = 0.89). The ESS original version reported Cronbach's alpha = 0.72 for students (ostensibly healthy people) [15], which is

lower than the value reported in the present study. Johns et al. [12] stated that about 10–20% of the general population have ESS scores >10. Interestingly, the prevalence of excessive daytime sleepiness in healthy Saudi subjects was found to be 19% at the cut-off point >10 in the initial test and 20% at the cut-off point >10 in the retest. These findings are in agreement with the report of Johns et al. [12] and are in contrast with the findings reported by Wali et al. [18].

On the first occasion (initial test), the value of Cronbach's alpha increased slightly to a value of 0.004 (e.g., Cronbach's alpha = 0.864) after deleting item number 6 ("sitting and talking to someone") from the survey. This, however, does not justify the deletion of this test question, as an increase of 0.004 is negligible. On the second occasion (retest), the deletion of any item from the survey did not improve the alpha value. Riachy et al. [16] present an Arabic version of ESS (ArESS) to investigate its reliability and validity, and Cronbach's alpha was 0.76.

As noted in this proposed study, the observed internal consistency was higher than the ArESS [16]. The discrepancy between their findings and these is likely to be due to the differences in the study design. Cultural differences inherited in the ESS and ArESS, such as gender-based prohibitions on the operation of motor vehicles and culture prohibitions of alcohol consumption were not accounted for in the former studies, perhaps contributing to the lower validity as compared to this study of ArESS. Further, the ArESS questionnaire was applied to patients with sleep-related breathing diseases and healthy people at one specific time, compared with this study which was only applied to healthy people at two different times for test reliability and consistency.

The present investigation of four-week test–retest reliability found that the AESS scores did not change significantly over time and respondents remain highly consistent with their initial survey compared with the retest. The mean difference of

**Table 2** Internal consistency and normal range of ArESS subscales.

ESS	Mean $\pm$ SD (95% CI for $\mu$ )	Cronbach's alpha (95% CI)	Range	Normal range (2.5–97.5 percentiles)
Test	6.3 $\pm$ 4.7 (5.330, 7.270)	0.86 (0.812, 0.900)	0–20	0–19.0
Retest	6.5 $\pm$ 5.3 (5.410, 7.590)	0.89 (0.852, 0.921)	0–20	0–21.7

**Table 3** Intra-class correlation of the ArESS sum scores.

Situation: test vs. retest	Paired differences		<i>t</i>	<i>P</i> -value	ICC(95% CI)
	Mean	SD			
Epworth Sleepiness Scale	–0.19	3.51	–0.51	0.611	0.86(0.789, 0.909)

ArESS scores between the initial test and retest was  $-0.19 \pm 3.51$ , which was not significantly different from 0 ( $p$ -value = 0.611). High reproducibility was observed on ArESS scores when the survey is applied twice to the same individuals (test–retest). The proposed study revealed that the intra-class correlation coefficient of test–retest was 0.86, which is considered almost perfect agreement and therefore supports the statement made by Johns that ESS is a very reliable instrument for measuring daytime sleepiness [15]. This study found the agreement of test–retest of the ArESS sum scores to be almost perfect, thereby supporting the usefulness of this instrument in assessing daytime sleepiness [16].

Similar to previous studies (e.g., Johns and Hocking, [12]), the current data from Saudi Arabia show that the ESS score does not differ significantly between males and females. In this study the average ESS score was ( $6.3 \pm 4.7$ , range 0–20), which is considered significantly higher than in Western cultures. For instance, the average ESS score in Australia [12] was ( $4.6 \pm 2.8$ , range 0–10,  $p$ -value = 0.005) and in Italy [17] it was ( $4.4 \pm 2.8$ , range 0–11,  $p$ -value = 0.003).

Inherent in this study are several limitations that should be noted. The effectiveness of the ArESS may be affected by cultural differences, especially

as seen in items 7 and 8. According to this study, females and males reported a similar likelihood of falling asleep in items 7 and 8. Further, there may be some benefit in adjusting the translation to account for cultural variation in meals. For example, lunch in Saudi Arabia may be considered the heaviest meal of the day, contributing to higher sleepiness than seen in a culture with less importance placed on lunch. This survey will be applied to a larger population to assess daytime sleepiness among patients with and without sleep disorders. Further research may be done to make the ArESS more culturally appropriate for Saudi Arabian audiences.

## 6. Competing interests

The authors declare that they have no competing interests.

## 7. Conflict of interest

None declared.

## Appendix A. Arabic Epworth Sleepiness

مقاس إيبورث: مافرصه حدوث إغفائه في اي من الحالات التاليه:				<i>Epworth Sleepiness Scale</i> Chance of dozing in any of the following situations	
عدد مرات الحدوث				الوضعية	Situation
دائما	غالبا	نادرا	ابدا		
3	2	1	0	خلال الجلوس والقراءة	Sitting and reading
3	2	1	0	خلال مشاهدة التلفاز	Watching TV
3	2	1	0	الجلوس هادئا في الاماكن العامة ( مثلا: في مسرح او غرف الانتظار او اثناء الاجتماعات)	Sitting inactive in a public place ((e.g. a theater or a meeting)
3	2	1	0	الجلوس كراكب في سفر لمدة ساعه من غير توقف	As a passenger in a car for an hour without a break
3	2	1	0	خلال التمدد للراحة بعد الظهر اذا سمحت الظروف	Lying down to rest in the afternoon when circumstances permit
3	2	1	0	خلال الجلوس والتحدث مع شخص اخر	Sitting and talking to someone
3	2	1	0	خلال الجلوس هادئا بعد الغداء	Sitting quietly after a lunch without alcohol
3	2	1	0	في السيارة اثناء التوقف نتيجة الزحام لعدة دقائق	In a car, while stopped for a few minutes in traffic
				المجموع	

## Scale (ArESS)

## References

- [1] Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep* 1994;17(8):703–10.
- [2] Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991;14(6):540–5.
- [3] Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth Sleepiness Scale. *Chest* 1993;103(1):30–6.
- [4] Izquierdo-Vicario Y, Ramos-Platon MJ, Conesa-Peraleja D, Lozano-Parra AB, Espinar-Sierra J. Epworth Sleepiness Scale in a sample of the Spanish population. *Sleep* 1997;20(8):676–7.
- [5] Bloch KE, Schoch OD, Zhang JN, Russi EW. German version of the Epworth Sleepiness Scale. *Respiration* 1999;66(5):440–7.
- [6] Chung KF. Use of the Epworth Sleepiness Scale in Chinese patients with obstructive sleep apnea and normal hospital employees. *J Psychosom Res* 2000;49(5):367–72.
- [7] Chen NH, Johns MW, Li HY, Chu CC, Liang SC, Shu YH, et al. Validation of a Chinese version of the Epworth sleepiness scale. *Qual Life Res* 2002;11(8):817–21.
- [8] Takegami M, Suzukamo Y, Wakita T, Noguchi H, Chin K, Kadotani H, et al. Development of a Japanese version of the Epworth Sleepiness Scale (JESS) based on item response theory. *Sleep Med* 2009;10(5):556–65.
- [9] Izci B, Ardic S, Firat H, Sahin A, Altinors M, Karacan I. Reliability and validity studies of the Turkish version of the Epworth Sleepiness Scale. *Sleep Breath* 2008;12(2):161–8.
- [10] Vignatelli L, Plazzi G, Barbato A, Ferini-Strambi L, Manni R, Pompei F, et al. Italian version of the Epworth sleepiness scale: external validity. *Neurol Sci* 2003;23(6):295–300.
- [11] Tsara V, Serasli E, Amfilochiou A, Constantinidis T, Christaki P. Greek version of the Epworth Sleepiness Scale. *Sleep Breath* 2004;8(2):91–5.
- [12] Johns MW, Hocking B. Daytime sleepiness and sleep habits of Australian workers. *Sleep* 1997;844–949.
- [13] Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
- [14] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86(2):420–8.
- [15] Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep* 1992;15(4):376–81.
- [16] Riachy M, Juvelikian G, Sleilaty G, Bazarbachi T, Khayat G, Mouradides C. Validation of the Arabic Version of the Epworth Sleepiness Scale: multicentre study. *Rev Mal Respir* 2012;29(5):697–704.
- [17] Manni R, Politini L, Ratti MT, Tartara A. Sleepiness in obstructive sleep apnea syndrome and simple snoring evaluated by the Epworth Sleepiness Scale. *J Sleep Res* 1999;8:319–20.
- [18] Wali SO, Mirdad S, Almobaireek A. Sleep disorders in Saudi Health care workers. *Ann Saudi Med* 1999;19(5):406–9.

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

**ScienceDirect**