Sleep Breath (2011) 15:377–384
DOI 10.1007/s11325-010-0343-6

ORIGINAL ARTICLE

The reliability and validity of the Korean version of the Epworth sleepiness scale

Yong Won Cho · Joo Hwa Lee · Hyo Kyung Son · Seung Hoon Lee · Chol Shin · Murray W. Johns

Received: 10 January 2010 / Revised: 23 February 2010 / Accepted: 24 February 2010 / Published online: 1 April 2010

Abstract
Objective The Epworth sleepiness scale (ESS) is widely used to measure the subject's average sleep propensity across those different situations in daily life, particularly in patients with sleep-disordered breathing. The purposes of this study were to test the hypothesis that the Korean version of the ESS (KESS) is valid and evaluate its usefulness.
Materials and methods We developed the KESS, which involved translating into Korean and then translating back into English to check its accuracy. A total of 273 participants (181 obstructive sleep apnea (OSA)—37 mild, 61 moderate, 83 severe, 32 simple snoring and 60 normal) were included in this study. All subjects completed the overnight polysomnograph and 53 of the total subjects were randomly selected for a retest with the questionnaire approximately 2~4 weeks later. The associations between KESS and the degree of OSA were examined through ANCOVA, adjusted for age, sex and BMI.
Results The total score and each item's score of KESS in patients with OSA were significantly higher than subjects with normal controls ($p<0.01$). As the severity of OSA increased, the KESS showed significantly increasing patterns (p for trend <0.01). The KESS in patient groups showed good internal consistency (Cronbach's $\alpha=0.90$) and test–retest reliability ($r=0.78$ to 0.93).
Conclusion The KESS is a reliable and valid tool for screening patients with daytime sleepiness in Korea.

Keywords Epworth sleepiness scale · Sleep-disordered breathing · Reliability · Validity

Y. W. Cho (✉) · J. H. Lee
Department of Neurology, Dongsan Medical Center,
Keimyung University School of Medicine,
194, Dongsan-dong, Choon-Gu,
Daegu, South Korea
e-mail: neurocho@dsmc.or.kr

H. K. Son
Department of Occupational & Environmental Medicine,
CHA Gumi Medical Center, CHA University,
Gumi, South Korea

S. H. Lee
Department of Otorhinolaryngology-Head and Neck Surgery,
Korea University,
Seoul, South Korea

C. Shin
Internal Medicine, College of Medicine, Korea University,
Seoul, South Korea

M. W. Johns
Sleep Disorders Unit, Epworth Hospital,
Melbourne, Australia

Introduction

Daytime sleepiness is a common complaint in those with sleep disorders [1, 2]. In Korea, 12.2% of the adult population has excessive daytime sleepiness and 4.5% in men and 3.2% in women have obstructive sleep apnea [3, 4]. Many people believe that the multiple sleep latency test (MSLT) is a gold standard method for measuring daytime sleepiness [5]. However, the MSLT is not used routinely, because it is expensive and patients need to stay in the sleep study room all day. The Epworth sleepiness scale (ESS) is widely used to measure the general level of daytime sleepiness, particularly among patients with sleep-disordered breathing [6, 7]. It is a convenient and

inexpensive evaluation method and has clinical significance ascertaining the subjective values of the patient's perception of sleepiness.

The ESS is comprised of eight questions, each asking about the subject's likelihood of dozing off or falling asleep in a particular situation that is commonly met in daily life. Thus, each ESS item score measures a particular "situational sleep propensity", and the sum of those item scores, i.e., the total ESS score, measures the subject's average sleep propensity across those different situations in daily life [8]. Respondents use a four-point scale for each of the eight questions. Higher scores indicate higher subjective sleepiness. The ESS has been translated into more than five different languages, and the validity and reliability in the respective languages have been studied [9–13]. The ESS came from a western country, so a direct translation may have some limitations for use in oriental countries, due to the cultural and economic differences. They need to test for validation and reliability before using it in each country.

The purpose of this study was to test the hypothesis that the Korean version of the ESS (KESS) is valid and evaluate its usefulness.

Materials and methods

We developed the Korean version of the ESS (KESS) after obtaining permission from the original developer. To prepare a Korean version of the questionnaire, the original ESS questionnaire was translated from English into Korean and then translated back into English by two sleep specialists and one linguist. We discussed the cultural and social differences with Dr. Murray Johns, the original developer, and concluded that we needed to adjust the questionnaire to be more suitable to the Korean environment. In question number four, we changed the word "car" to "vehicle", which would include buses and trains. And we added one more situation: while sitting and writing a letter or document, because we expect that question number eight may have a lower response rate in Korea. Together with the original developer, we compared the original and the translated version in creating the preliminary Korean version. Each interview was conducted individually by a trained interviewer. The interviewer asked whether they regularly encountered the situations described in each question and whether they could imagine themselves in those situations even if they do not regularly encounter them. This step helped to reduce the percentage of missing data. When we did the pilot study with nine questions, the response rate to the original question number eight was not low. The response to question number nine was similar to those of the original question number eight. So, finally we used the original questionnaire with only one adjustment to question number four as follows: as a passenger in a vehicle (for example: a car, a bus, or a train) for an hour without a break (Appendix A and B).

The study was approved by the Institutional Review Board of Keimyung University, School of Medicine.

Subjects

A total of 273 participants (181 obstructive sleep apnea (OSA), 32 simple snoring and 60 normal healthy subjects) were included in this study. The subjects were from two university hospitals in different cities, and there were no significant differences in demographic factors.

A diagnostic confirmation of all the subjects was done by the overnight polysomnography (PSG). Using the apnea–hypopnea index (AHI), 181 subjects with OSA were divided into three groups: the mild group was made of subjects with AHI <15/h, n=37; the moderate group was comprised of those with AHI between 15 and 30/h, n=61; the severe group was comprised of those with AHI ≥30/h, n=83. And, if the subjects with a history of snoring had less than five AHI, we categorized them as simple snoring. Subjects with the other sleep disorders or with major systemic comorbidities or psychiatric conditions were excluded. For the healthy subjects, there were no sleep complaints, which were also confirmed by the PSG. All of the subjects completed the questionnaires, which included the KESS and the other sleep questionnaires, such as the Korean version of the Pittsburg Sleep Quality Index (PSQI). This survey was conducted by well-trained interviewers in the evening before the PSG study. To examine the test–retest reliability, 25% each of the five groups were randomly selected. About 2% of these subjects refused to retest or had treatment for OSA. Then, 53 of the total subjects were retested with the questionnaire approximately 2~4 weeks later.

A full PSG was performed on all subjects. The sleep stages were scored according to the Rechtschaffen and Kales criteria [14]. Obstructive apnea episodes were defined by cessation of airflow for more than 10 s in the presence of respiratory effort. Obstructive hypopnea was defined as a decrease in airflow by more than 50% in the presence of respiratory effort, with >3% desaturation and/or electroencephalogram arousal. The AHI was defined as the number of obstructive apnea and hypopnea events per hour [15].

Statistical analysis

The patients with OSA were categorized in three groups using severity of OSA. The associations between KESS and the degree of OSA were examined through ANCOVA, adjusted for age, sex and BMI. The internal consistency was tested by means of Cronbach's α, which is based on correlations of

items on a single scale. Test–retest reliability was evaluated by means of intraclass correlation coefficients.

All of the analyses were performed by using SPSS 14.0 for Windows (SPSS, Inc., Chicago, IL, USA). A probability of $p<0.05$ was considered statistically significant.

Results

KESS scores in normal subjects and in various patient groups

The demographic characteristics and PSG findings of the 181 patients divided into various diagnostic categories and healthy controls are listed in Table 1. All of the patients suffered from sleep-related breathing disorders. The mean KESS score of the healthy controls (5.07±2.93) was significantly lower than that in the 213 patients (8.21± 4.23). The BMI of subjects with moderate and severe apnea was significantly different than those of the healthy controls ($p<0.01$). In the PSG findings, the severe apnea group had less slow-wave sleep and more light sleep than the healthy controls ($p<0.01$) (Table 1).

Validation of the KESS

Compared with controls, total scores and each item's score of KESS were significantly higher in patients ($p<0.01$) (Fig. 1). KESS scores showed a significantly increased pattern with each group from healthy control to severe OSA; these results were found even after adjusting for age, sex and BMI (p for trend <0.01) (Table 2). And, as the scores from daytime dysfunction domain of the PSQI scores increased, the KESS showed significantly increasing patterns (p for trend <0.01). The KESS total score was 4.5 for the best daytime function and 10.4 for the group with the worst daytime function.

Internal consistency and test–retest reliability of the KESS

In patients, the mean score of each item showed a relatively wide range from 0.69 in item 6 to 1.98 in item 5. However, the correlation coefficients of the item with the total of scores on all other items were relatively strong (0.51–0.67). Cronbach's α coefficient for total score of KESS was 0.90. And, there was no increase in Cronbach's α if any of the items were deleted (Table 3).

Fifty three subjects were retested with the questionnaires approximately 2~4 weeks later. The KESS showed good internal consistency and test–retest reliability. The test–retest intraclass correlation coefficients were from 0.78 to 0.93. Reproducibility was good. No significant differences were found in any of the items or in the total score in the first and second assessments ($p>0.05$) (data not shown). The Spearman correlation coefficient between test and retest was $r=0.86$ ($p<0.001$).

Table 1 Demographic data, PSG findings and KESS scores by groups

	Healthy controls (n=60)	Simple snoring (n=32)	Obstructive sleep apnea		
			Mild apnea (n=37)	Moderate apnea (n=61)	Severe apnea (n=83)
Age (years)	43.75±13.87	36.34±14.94	45.57±11.85	46.85±13.52*****	43.55±9.31
Gender (M/W)	45/15	22/10	28/9	49/11	80/3
BMI (kg/m^2)	24.33±4.93	23.55±2.87	25.23±3.25	26.43±3.01***·*****	27.29±2.79*·****
PSG findings					
Sleep efficiency	91.89±3.83	83.66±10.10	80.16±16.14**	84.39±14.24***	82.67±14.79**
Light sleep (%)	63.17±25.64	61.09±10.93	62.54±15.17	76.70±47.13	77.57±13.12***
Slow-wave sleep (%)	19.09±10.07	19.78±9.43	18.17±13.20	11.23±11.73**	6.04±9.15*
REM sleep (%)	22.72±12.60	18.91±5.34	17.37±7.14	16.65±6.99**	14.37±7.61*
KESS scores	5.07±2.93	7.50±4.16	7.92±3.37***	7.62±3.19***	9.01±5.10*

Mean±SD

PSG polysomnography, *KESS* Korean version of the Epworth sleepiness scale, *M* men, *W* women, *BMI* body mass index, *REM* rapid eye movement

*$p<0.001$ vs. healthy controls

**$p<0.01$ vs. healthy controls

***$p<0.05$ vs. healthy controls

****$p<0.001$ vs. simple snoring

*****$p<0.01$ vs. simple snoring

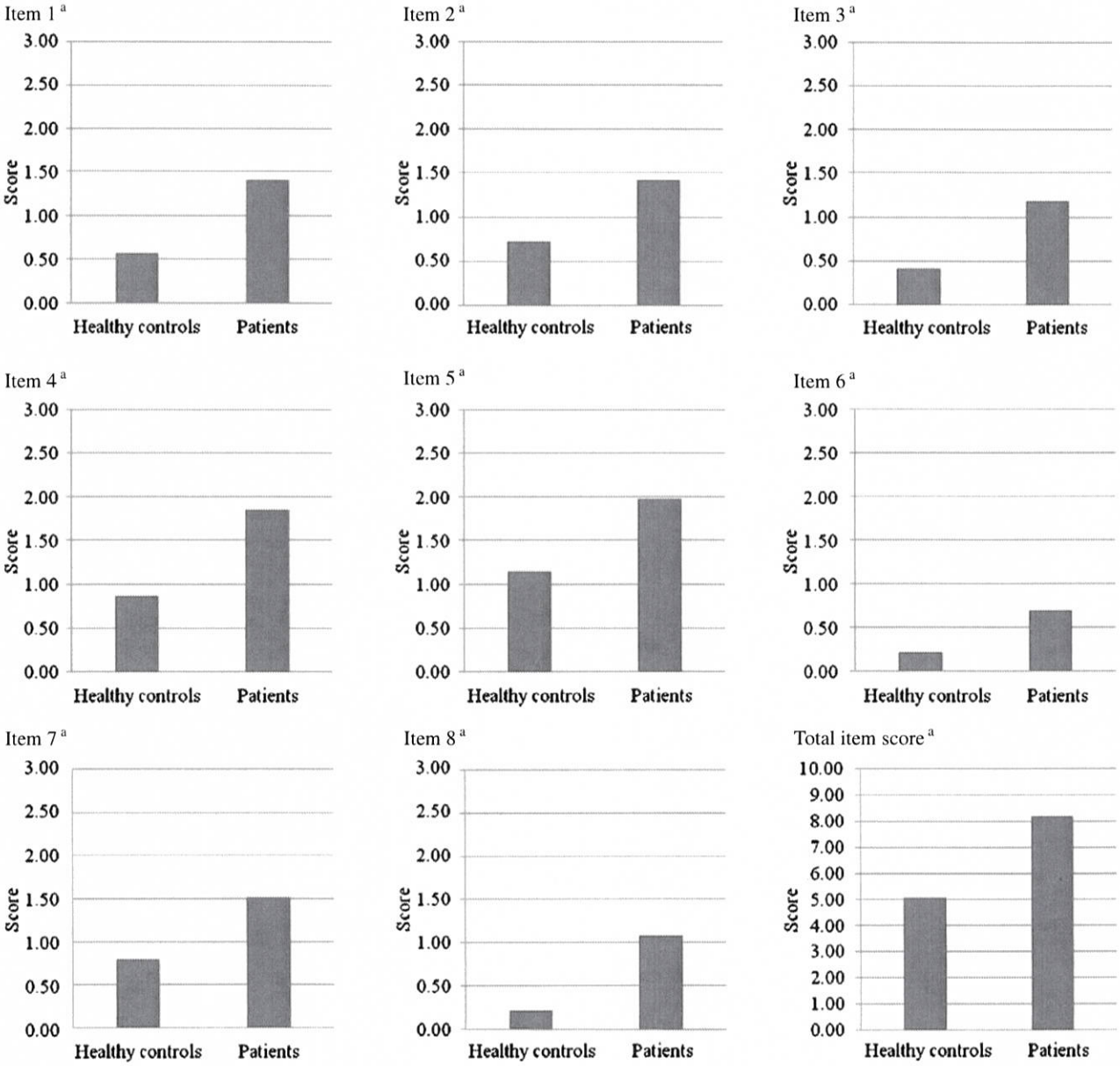

Fig. 1 Comparisons of ESS scores of patients with OSA and healthy controls. [a]$p<0.001$. *Item 1* sitting and reading; *Item 2* watching TV; *Item 3* sitting, inactive in a public place; *Item 4* as a passenger in a vehicle for an hour without a break; *Item 5* lying down to rest in the afternoon when circumstances permit; *Item 6* sitting and talking to someone; *Item 7* sitting quietly after a lunch without alcohol; *Item 8* in a car, while stopped for a few minutes in the traffic

Discussion

This study showed that the KESS measures the sleepiness of subjects with OSA similar to that measured by the original ESS. To reflect routine daily activity of Koreans, we changed question number four of the ESS and added one more question to be more suitable to Korean culture. After the pilot study with these nine questions, we developed the KESS, which has only one adjustment to question number four as follows: as a passenger in a vehicle (for example: a car, a bus, or a train) for an hour without a break.

The total ESS score gives a measure of the subject's average sleep propensity in daily life. None of the other commonly used methods for measuring "sleepiness", such as the MSLT, attempt to do that [16, 17]. For the validation study, we did the questionnaire in healthy people and sleep-disordered breathing subjects. All of the subjects were confirmed by a full PSG. The mean KESS scores and each

Table 2 Adjusted KESS scores in patients with sleep breathing disorders and in healthy controls

	Healthy controls (n=60)	Simple snoring (n=32)	Obstructive sleep apnea			p for trend[b]
			Mild apnea (n=37)	Moderate apnea (n=61)	Severe apnea (n=83)	
Sitting and reading	0.59 (0.34–0.85)[a]	1.00 (0.68–1.32)	1.19 (0.89–1.50)	1.31 (1.06–1.55)	1.41 (1.17–1.65)	<0.001
Watching TV	0.85 (0.60–1.11)	1.02 (0.70–1.34)	1.39 (1.09–1.69)	1.30 (1.05–1.54)	1.52 (1.28–1.76)	<0.001
Sitting, inactive in a public place	0.46 (0.23–0.69)	0.97 (0.68–1.26)	1.11 (0.84–1.38)	1.01 (0.79–1.24)	1.19 (0.97–1.41)	<0.001
As a passenger in a vehicle for an hour without a break	0.96 (0.65–1.23)	1.45 (1.08–1.81)	1.84 (1.50–2.19)	1.87 (1.58–2.15)	1.83 (1.55–2.10)	<0.001
Lying down to rest in the afternoon when circumstances permit	1.29 (1.01–1.56)	1.45 (1.10–1.79)	1.78 (1.46–2.10)	2.06 (1.80–2.33)	2.18 (1.93–2.44)	<0.001
Sitting and talking to someone	0.28 (0.09–0.46)	0.51 (0.27–0.74)	0.52 (0.30–0.73)	0.73 (0.55–0.91)	0.75 (0.58–0.93)	<0.001
Sitting quietly after a lunch without alcohol	0.83 (0.58–1.08)	1.21 (0.89–1.52)	1.28 (0.99–1.57)	1.36 (1.12–1.60)	1.58 (1.35–1.82)	<0.001
In a car, while stopped for a few minutes in the traffic	0.25 (0.00–0.51)	0.78 (0.46–1.10)	0.79 (0.50–1.09)	1.09 (0.84–1.33)	1.27 (1.03–1.51)	<0.001
Total KESS score	5.38 (4.25–6.50)	7.53 (6.11–8.94)	7.77 (6.46–9.09)	7.15 (6.07–8.23)	7.90 (6.85–8.95)	0.003

KESS Korean version of the Epworth sleepiness scale

[a] Values in parentheses correspond to 95% confidence intervals

[b] Significant test: ANCOVA, means adjusted for age, sex and BMI

item's score of KESS have discriminate validity between normal controls and subjects with sleep-disordered breathing. The KESS scores showed significantly increasing patterns with each group from normal people to severe OSA, even though the total KESS scores of the moderate OSA group were not higher than the mild OSA group. There were no statistically significant differences between mild and moderate subjects with OSA. This result suggests that the KESS distinguishes between normal controls and subjects with sleep-disordered breathing, but is not sensitive enough to distinguish between each of the OSA subgroups. Some previous studies reported that the ESS was not able to differentiate the degree of sleepiness in relation to the severity of sleep-disordered breathing and showed no correlation between ESS scores and AHI [18, 19]. In our study, the mean unadjusted KESS score and adjusted KESS score of severe OSA subjects were 9.01 and 7.93, respectively. This score was relatively lower than in other countries [11, 13, 20] because, in our population, some subjects with OSA did not complain about excessive daytime sleepiness and also due to the difference in the tendency of snorers to seek medical attention in each country. Some previous studies also reported that there was no relationship between daytime sleepiness and the severity of sleep-disordered breathing even when an objective method, such as MSLT, was used [21, 22].

The KESS showed significantly increasing patterns as the scores from the daytime dysfunction domain of the PSQI scores increased. This result indicates that the mean score of the KESS is related with the daytime dysfunction

Table 3 Item analysis of the KESS in patients with sleep breathing disorders and in healthy controls

Item number	Patients (n=213)		Healthy controls (n=60)	
	Item to total correlation	Cronbach's α[a]	Item to total correlation	Cronbach's α[a]
1	0.59	0.89	0.63	0.69
2	0.51	0.89	0.46	0.74
3	0.66	0.89	0.62	0.69
4	0.67	0.89	0.71	0.67
5	0.58	0.89	0.63	0.69
6	0.53	0.89	0.66	0.68
7	0.62	0.89	0.53	0.71
8	0.63	0.89	0.48	0.72
Total		0.90		0.73

KESS Korean version of the Epworth sleepiness scale

[a] Cronbach's alpha, if the item was deleted

domain of the PSQI scores. The daytime function domain of the PSQI consists of alertness and enthusiasm, so the daytime dysfunction score also represents some degree of sleep propensity.

The KESS showed good internal consistency and test–retest reliability, comparable to those of other translations of the ESS [9, 10, 20]. Item analysis of the KESS revealed results that were very close to those reported for the original version. In normal subjects, the scores for question number six and eight were very low and consistent with the low-factor loading of these questions in the original version [23]. The internal consistency of the questionnaire in healthy controls was fair with a Cronbach's α value of 0.73. Patients, Cronbach's α of 0.90, higher than the previously reported value of 0.88, indicated a high level of internal consistency.

This study has several limitations. First, it would be better to compare total scores of the KESS with objective measured values by the MSLT. However, MSLT is an expensive and time-consuming method. Also, the situation of MSLT is not the same in routine daily activity environments. Second, in our study, all of the subjects were healthy subjects or patients with sleep-disordered breathing. Further studies are needed for the other sleep disorders, such as narcolepsy, idiopathic hypersomnia, insomnia, etc. Third, the BMI and the number of male subjects were significantly higher for subjects with sleep-disordered breathing. That was the expected results and it was not easy to find healthy male subjects with a high BMI. However, our subjects are from two different and separate locations, and there were no significant differences in demographic factors. Therefore, we believe our data to be representative of South Koreans.

In conclusion, the pattern between the scores of KESS in healthy controls and patients with sleep breathing disorders is similar to those of the original English ESS. It is a reliable and valid tool. Therefore, it is useful for the measurement of the subject's average sleep propensity in patients with sleep disorders in the South Korean population.

Acknowledgements This work was supported by the research promoting grant from the Keimyung University Dongsan Medical Center in 2009.

Appendix 1. The Korean version of the Epworth sleepiness scale

A. Korean version of the Epworth sleepiness scale (English)

Name : _____

Today's date : _____ Your age (in years) : _____

Your sex (male = M; female = F) : _____

How likely are you to doze off or fall asleep in following situations, in contrast to feeling just tired? This refers to your usual way of life in recent times. Even if you have not done some of these things recently try to work out how day would have affected you. Use of following scale to choose the most appropriate number for each situation:

0=would *never* doze
1=*slight* chance of dozing
2=*moderate* chance of dozing
3=*high* chance of dozing

Situation	Chance of dozing
Sitting and reading	_____
Watching TV	_____
Sitting, inactive in a public place (e.g. a theater or a meeting)	_____
As a passenger in vehicles (e.g. a car, a bus, or a train) for an hour without a break	_____
Lying down to rest in the afternoon when circumstances permit	_____
Sitting and talking to someone	_____
Sitting quietly after a lunch without alcohol	_____
In a car, while stopped for a few minutes in the traffic	_____

Thank you for your cooperation

B. Korean version of the Epworth sleepiness scale (Korean)

한국형 주간졸음척도 (KESS)

단순한 피곤함과는 다르게 다음의 상황에서 얼마나 깜박 졸거나 혹은 잠들어버릴 것 같습니까? 이것은 최근에 **일상생활을 참고**하시기 바랍니다. 비록 최근에 이런 상황에 처하지 않았다 하더라도, 그 상황에서 얼마나 영향을 받을지 생각해 보십시오. 각 상황에서 **가장 적절한** 숫자를 선택하여 **한 개씩만 √표**하시기 바랍니다.	깜박 졸 가능성			
	전혀 0	조금 1	상당히 2	매우 많이 3
앉아서 책 (신문, 잡지, 서류 등)을 읽을 때				
TV 볼 때				
공공장소 (모임, 극장 등)에서 가만히 앉아 있을 때				
정차 없이 1시간 동안 운행 중인 차 (자동차, 버스, 열차)에 승객으로 앉아 있을 때				
오후에 주위상황이 허락되어 쉬려고 누워 있을 때				
앉아서 상대방과 이야기할 때				
반주를 곁들이지 않은 점심식사 후 조용히 앉아 있을 때				
교통 혼잡으로 몇 분 동안 멈춰선 차 안에서				

References

1. Baran AS, Chervin RD (2009) Approach to the patient with sleep complaints. Semin Neurol 29(4):297–304
2. Schwartz JR, Roth T, Hirshkowitz M, Wright KP (2009) Recognition and management of excessive sleepiness in the primary care setting. Prim Care Companion J Clin Psychiatry 11 (5):197–204
3. Joo S, Baik I, Yi H, Jung K, Kim J, Shin C (2009) Prevalence of excessive daytime sleepiness and associated factors in the adult population of Korea. Sleep Med 10(2):182–188
4. Kim J, In K, You S, Kang K, Shim J, Lee S, Lee J, Park C, Shin C (2004) Prevalence of sleep-disordered breathing in middle-aged Korean men and women. Am J Respir Crit Care Med 170 (10):1108–1113
5. Wise MS (2006) Objective measures of sleepiness and wakefulness: application to the real world? J Clin Neurophysiol 23(1):39–49
6. Johns MW (1993) Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth sleepiness scale. Chest 103(1):30–36
7. Johns MW (1991) A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 14(6):540–545
8. Johns MW (1994) Sleepiness in different situations measured by the Epworth Sleepiness Scale. Sleep 17(8):703–710
9. Beiske KK, Kjelsberg FN, Ruud EA, Stavem K (2009) Reliability and validity of a Norwegian version of the Epworth sleepiness scale. Sleep Breath 13(1):65–72
10. Bloch KE, Schoch OD, Zhang JN, Russi EW (1999) German version of the Epworth sleepiness scale. Respiration 66(5):440–447
11. Izci B, Ardic S, Firat H, Sahin A, Altinors M, Karacan I (2008) Reliability and validity studies of the Turkish version of the Epworth sleepiness scale. Sleep Breath 12(2):161–168
12. Takegami M, Suzukamo Y, Wakita T, Noguchi H, Chin K, Kadotani H, Inoue Y, Oka Y, Nakamura T, Green J, Johns MW, Fukuhara S (2009) Development of a Japanese version of the Epworth Sleepiness Scale (JESS) based on item response theory. Sleep Med 10(5):556–565
13. Tsara V, Serasli E, Amfilochiou A, Constantinidis T, Christaki P (2004) Greek version of the Epworth sleepiness scale. Sleep Breath 8(2):91–95
14. Rechtschaffen A, Kales A (1973) A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. National Institute of Health (USA) publication no. 204. Brain Information Service/Brain Research Institute, University of California, Los Angeles, California
15. Force AAoSMT (1999) Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. Sleep 22(5):667–689

16. Johns MW (2000) Sensitivity and specificity of the multiple sleep latency test (MSLT), the maintenance of wakefulness test and the Epworth sleepiness scale: failure of the MSLT as a gold standard. J Sleep Res 9(1):5–11
17. Johns MW (2002) Sleep propensity varies with behaviour and the situation in which it is measured: the concept of somnificity. J Sleep Res 11(1):61–67
18. Osman EZ, Osborne J, Hill PD, Lee BW (1999) The Epworth sleepiness scale: can it be used for sleep apnoea screening among snorers? Clin Otolaryngol Allied Sci 24(3):239–241
19. Chervin RD, Aldrich MS (1999) The Epworth sleepiness scale may not reflect objective measures of sleepiness or sleep apnea. Neurology 52(1):125–131
20. Chen NH, Johns MW, Li HY, Chu CC, Liang SC, Shu YH, Chuang ML, Wang PC (2002) Validation of a Chinese version of the Epworth sleepiness scale. Qual Life Res 11 (8):817–821
21. Chervin RD (2000) The multiple sleep latency test and Epworth sleepiness scale in the assessment of daytime sleepiness. J Sleep Res 9(4):399–401
22. Furuta H, Kaneda R, Kosaka K, Arai H, Sano J, Koshino Y (1999) Epworth sleepiness scale and sleep studies in patients with obstructive sleep apnea syndrome. Psychiatry Clin Neurosci 53 (2):301–302
23. Johns MW (1992) Reliability and factor analysis of the Epworth sleepiness scale. Sleep 15(4):376–381