Quality of Life Research **11**: 817–821, 2002.

Validation of a Chinese version of the Epworth sleepiness scale

Ning-Hung Chen[1], Murray W. Johns[2], Hseuh-Yu Li[3], Chia-Chen Chu[4], Shu-Cheng Liang[4], Yu-Hsiang Shu[5], Ming-Lung Chuang[1] & Pa-Chun Wang[6,7]

[1]*Department of Pulmonary and Critical Care Medicine, Sleep Center, Chang Gung Memorial Hospital (E-mail: drtony@tpts4.seed.net.tw);* [2]*Sleep Disorders Unit, Epworth Hospital, Melbourne, Australia;* [3]*Department of Otolaryngology, Chang Gung Memorial Hospital and Chang Gung University, Taipei, Taiwan;* [4]*Department of Health, Taipei, Taiwan;* [5]*Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan;* [6]*Department of Otolaryngology, Cathay General Hospital;* [7]*Department of Public Health, China Medical College, Taichung Taiwan, ROC*

Accepted in revised form 30 April 2002

Abstract

Epworth sleepiness scale (ESS) is widely used to evaluate degree of somnolence among Chinese patients with sleep-disordered breathing. Yet no Mandarin–Chinese translation has ever reported its validation data. In this study we translated and validated the ESS into Mandarin Chinese (CESS). We found, in 31 bilingual patients' responses to the CESS and the English ESS obtained 1–2 weeks apart did not differ significantly (10.5 ± 3.7 vs. 9.6 ± 3.9, $p = 0.32$, Wilcoxon's signed rank test) and were significantly correlated (Spearman's $\rho = 0.67$, $p = 0.0004$). A total of 359 sleep-disordered breathing subjects were enrolled into the validation study in a prospective manner. The CESS showed acceptable internal consistency (Cronbach's $\alpha = 0.81$). Thirty out of these patients answered CESS twice at an interval of 2–4 weeks, to yield an acceptable level of test–retest reliability ($\rho = 0.74$, $p = 0.001$). The respiratory disturbance indices obtained from 251 out of 359 subjects were significantly correlated with their CESS scores ($\rho = 0.22$, $p < 0.001$). The standard response mean of CESS was 0.86 as obtained from 94 out of 359 subjects who had initial control of their symptoms at 3 months after radio-frequency palate surgery. We conclude that the CESS is reliable in both a linguistic and a test–retest sense, and appears to be valid and sensitive to clinical change. The CESS could be used to advantage among other Mandarin-speaking subjects as a standardised screening test of sleepiness in daily life.

Key words: Epworth, Mandarin–Chinese, Sleepiness

Background

Sleep-disordered breathing covers a wide spectrum of breathing disorders during sleep including snoring, upper airway resistance syndrome, obstructive sleep apnea syndrom (OSAS), and obese hypoventilation. Sleep-disordered breathing and its sequelae are major quality-of-life issue [1].

Overnight polysomnography (PSG) is the standard test for the diagnosis of sleep-disordered breathing. The multiple sleep latency test (MSLT) has been used as the gold standard [2, 3] to evaluate daytime sleepiness. However, the cost-effectiveness of the MSLT as an accurate quantitative test of sleepiness in daily life has been challenged [4]. The development of multiple treatment modalities for OSAS [5–7] has further induced demand for more efficient evaluation tools. Uses of self-reported health surveys to screen patients with OSAS have been proposed [8–10].

The 8-item Epworth sleepiness scale (ESS) (Appendix 1) is a useful tool for evaluating adults

of the average sleep propensity in daily life [11]. Each of the 8-item–scores can range from 0 to 3 and the total Epworth score from 0 to 24 (lowest to highest sleep propensity). The reliability, unitary structure and validity of the ESS is supported by a rapidly growing body of experimental evidence in distinguishing the excessive daytime sleepiness of narcoleptics from that of normal subjects [4, 12], or in quantifying the reduction in sleepiness of OSAS patients after treatment with nasal CPAP [13].

To maintain its usefulness and to allow comparisons between results from different centres, it is important that the ESS be standardised, particularly when it is translated into another language. Currently, there is no validated translation of the ESS into Chinese (Mandarin). We conducted a study to translate and validate the CESS, also to test its use among Chinese patients suffering from sleep-disordered breathing.

Translation of ESS into Mandarin Chinese

Translation of the English ESS into Chinese (CESS) (Appendix 2) followed the standard forward-step, backward-step, and pretest-step method [14]. The ESS was first translated into Mandarin Chinese and then back into English iteratively by two bilingual physicians other than the translator (Chen, NH) until both versions were considered completely interchangeable, conceptually and linguistically. The pre-test step involved a bilingual lay panel of three people who assessed the comprehensibility of the CESS and tested translation alternatives, highlighting unexpected errors and inappropriate terms. A separate panel of 31 bilingual testees with documented sleep-disordered breathing were randomly selected to answer firstly the ESS and then CESS at 1–2 week intervals. Their scores on each version were as one measure of success with the translation.

We found that Epworth scores from the final version of the CESS and the English ESS that were both answered by the panel of 31 bilingual subjects did not differ significantly (10.5 ± 3.7 vs. 9.6 ± 3.9, $p = 0.32$, Wilcoxon signed rank test) and were significantly correlated ($\rho = 0.67$; $p = 0.0004$). This indicates very good linguistic interchangeability between the translation and the original questionnaire. This data is also comparable to the correlation between the test and retest Epworth scores of 30 sleep-disordered breathing patients who answered the CESS twice ($\rho = 0.74$; $p = 0.001$) in our validation study.

Reliability of CESS

A total of 359 patients who visited our clinic with complaints of daytime sleepiness and symptoms of sleep-disordered breathing were consecutively enrolled into our study in a prospective manner. A Spearman's correlation level of 0.5 was considered acceptable in this study. Thirty patients from this cohort answered the translated CESS twice with a 2–4 weeks interval. The test–retest reliability of the individual CESS item–scores are shown in Table 1. All correlations except for item-6 were statistically significant. The mean of those Spearman's correlation coefficients was 0.55 which represents an acceptable reliability level for CESS item–scores and was also very similar to that reported previously (mean $\rho = 0.56$) [15]. The test–retest reliability of CESS ($\rho = 0.74$) is similar to that of ESS (0.81, $p < 0.001$) [15], and is also comparable with those of many MSLT studies (0.65–0.97) [16–19].

The internal consistency of the CESS for all 359 patients, measured by Cronbach's statistic, α, was acceptable = 0.81 (>0.7 is considered acceptable) and was very similar to that reported previously (0.74–0.88) [15, 20]. Spearman's ρ between scores for each CESS item and the total scores varied between 0.43 (item-6) and 0.77 (item-5) (Table 1). All were statistically significant ($p < 0.01$). Factor analysis of the item–scores for 359 patients using Kendall's coefficient of concordance yielded only one factor with an eigen value = 3.03, with normalised factor loadings >0.55 for all items except item-6 for which the loading was = 0.4 (Table 1). These results are also comparable to others published previously in relation to English-speaking patients [15, 21, 22].

Validity of CESS

A separate sex, age and body mass index-matched group of 31 ostensibly healthy adults who did not snore or have complaints of insomnia or poor

Table 1. The mean ± SD CESS item–scores for 359 patients, their test–retest correlations for 30 patients, the item vs. total CESS correlations with their statistical significance (Spearman's ρ), and the normalised factor loadings from factor analysis of CESS item–scores in 359 patients

Item no.	Item–scores mean ± SD	Item–scores test–retest		Item vs. total scores		Normalised factor loadings
		Correlation (ρ)	p	Correlation (ρ)	p	
1. Sitting and reading	1.40 ± 1.04	0.86	0.0001	0.62	0.0005	0.63
2. Watching TV	1.29 ± 0.92	0.62	0.0005	0.62	0.0005	0.67
3. Sitting inactive in a public place (e.g. a theatre or a meeting)	1.30 ± 0.97	0.52	0.005	0.69	0.0001	0.66
4. As a passenger in a car for an hour without a break	1.71 ± 1.00	0.50	0.008	0.75	0.0001	0.63
5. Lying down to rest in the afternoon when circumstances permit	2.27 ± 0.80	0.39	0.046	0.77	0.0001	0.68
6. Sitting and talking to someone	0.57 ± 0.77	0.23	>0.05	0.43	0.02	0.36
7. Sitting quietly after a lunch without alcohol	1.55 ± 1.02	0.77	0.001	0.76	0.0001	0.68
8. In a car while stopped for a few minutes in the traffic	0.69 ± 0.85	0.48	0.011	0.56	0.0002	0.56
Number of subjects	359	30		359		359

quality sleep were used to answer the CESS as representatives of 'normal' sleepers.

The Epworth scores for all 359 patients were 11.3 ± 5.6 (mean ± standard deviation), with a range from 0 to 24. Accepting that the variables were not all normally distributed, non-parametric methods were used and statistical significance was accepted at $p < 0.05$ in two-tailed tests. These scores are significantly higher than those of the 31 'normal' sleepers, for which the mean was 7.1 ± 4.5 (Mann–Whitney U-test, $p = 0.0001$).

The complete results of PSG were available for 251 out of our initial 359 sleep-disordered breathing patients. Patients with major systemic co-morbidity such as hypertension, diabetes mellitus, chronic obstructive airways disease or cardiac failure were excluded. The mean age of the 251 patients was 44.8 ± 11.0 year (range 24–63, male 53.8%). Their mean body mass index was 26.3 ± 4.8. The respiratory disturbance index (RDI, defined as the number of apneas and hypopneas, each lasting at least 10 sec, per hour of sleep) was 26.7 ± 24.6. Their RDIs were used as a measure of the severity of sleep-disordered breathing and were correlated with total CESS scores to provide evidence for the validity of the CESS. The correlation between RDI and CESS was significant (Spearman's $\rho = 0.22$, $p < 0.001$), but this was not a close relationship.

The relationship between daytime sleepiness and the severity of sleep-disordered breathing, no matter how those parameters have been measured, has never been reported as a close one. In this respect, ESS or CESS scores are no different from the results of much more expensive and time-consuming MSLTs [3]. The reality is that there is only a weak relationship between daytime sleepiness and severity of sleep-disordered breathing [13].

Responsiveness of CESS

A total of 94 out of these 359 patients had radiofrequency palatoplasty. All reported initial control of their sleep-disordered breathing symptoms. The mean of Epworth score was 10.1 ± 6.1 before surgery, 6.6 ± 4.7 at 3 months after surgery. The improvement was significant (Wilcoxon's signed rank test, $p = 0.0007$) and the ESS score of less than 10 was considered non-pathological clinically. Longitudinal sensitivity of the CESS to clinical change was calculated as the standard response mean (SRM = response mean/response standard deviation) described by Liang et al. [23]. According to established criteria, the SRM of more than 0.8 usually indicates an excellent responsiveness; the SRM for the CESS was high (0.86) based on this standard.

Appendix 1. The Epworth sleepiness scale

Epworth sleepiness scale
How likely are you to doze off or fall asleep in the 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 they would have affected you. Use the following scale to choose the *most appropriate number* for each situation:
0 – would *never* doze; 1 – *slight* chance of dozing; 2 – *moderate* change of dozing; 3 – *high* chance of dozing.

1. Sitting and reading	0	1	2	3
2. Watching TV	0	1	2	3
3. Sitting, inactive in a public place (e.g. a theater or a meeting)	0	1	2	3
4. As a passenger in a car for an hour without a break	0	1	2	3
5. Lying down to rest in the afternoon when circumstances permit	0	1	2	3
6. Sitting and talking to someone	0	1	2	3
7. Sitting quietly after a lunch without alcohol	0	1	2	3
8. In a car, while stopped for a few minutes in the traffic	0	1	2	3

Appendix 2. Mandarin Chinese version Epworth sleepiness scale

Epworth嗜睡問卷調查

請圈選出在以下不同情況中您打瞌睡的頻率:
0:從未, 1:很少, 2:一半以上 3:幾乎都會

1. 坐著閱讀時.	0	1	2	3
2. 看電視時	0	1	2	3
3. 在公眾場合安靜坐著(如在戲院或會議中)	0	1	2	3
4. 坐車連續超過一小時(不包含自己開車)	0	1	2	3
5. 在下午躺下休息時	0	1	2	3
6. 坐著與人交談時	0	1	2	3
7. 沒有喝酒的情況下在午餐後安靜坐著時	0	1	2	3
8. 開車中遇到交通問題而停下數分鐘時	0	1	2	3

The changes in sleepiness in this investigation are comparable to those reported previously after the treatment of sleep-disordered breathing by nasal CPAP [13]. However, high SRM does not necessarily indicate a very sensitive survey, that can also be in part ascribed to the significant and consistent improvement in CESS after radio-frequency treatment. Limitation in using the SRM as a measure of responsiveness should be noted; nevertheless, the CESS proved to be able to capture this clinical change.

Conclusion

Our translation of the ESS into Chinese appears to be accurate, reliable and valid. In summary, we have presented statistics evidences that our CESS translation is accurate in a linguistic sense, reliable and valid. It has psychometric features that make it suitable for others to use as a standardised test of sleepiness for screening or clinical trial among the very many Mandarin-speaking people.

Acknowledgements

Research was sponsored by the Chang Gung Memorial Hospital Medical Research Grant 0986.

References

1. Kryger M, Roth T, Dement W. Principles and Practice of Sleep Medicine, 3rd edn. Philadelphia: W.B. Sanders, 2000.
2. Thorpy MT. The clinical use of the multiple sleep latency test. Sleep 1992; 15: 268–276.

3. Chervin RD, Aldrich MS. The Epworth sleepiness scale may not reflect objective measures of sleepiness or sleep apnea. Neurology 1999; 52: 125–131.
4. Johns MW. 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 2000; 9: 5–11.
5. Fujita A, Conway W, Zorick F, et al. Surgical correction of anatomic abnormalities of obstructive sleep apnea syndrome: Uvulopalatopharyngoplasty. Otolaryngol Head Neck Surg 1981; 106: 10–13.
6. Powell NB, Riley RW, Guilleminault C. Radiofrequency tongue base reduction in sleep-disordered breathing: A pilot study. Otolaryngol Head Neck Surg 1999; 120: 656–664.
7. Li KK, Powell NB, Riley RW, et al. Radiofrequency volumetric reduction of the palate. Otolaryngol Head Neck Surg 2000a; 122: 410–414.
8. Li KK, Riley RW, Powell NB, Gervacio L, Troell RJ, Guilleminault C. Obstructive sleep apnea surgery: Patient perspective and polysomnogram results. Otolaryngol Head Neck Surg 2000b; 123: 572–575.
9. Pradhan PS, Gliklich RE, Winkelman J. Screening for obstructive sleep apnea in patients presenting for surgery. Laryngoscope 1996; 106: 1393–1397.
10. Simmons FB, Guilleminault C, Miles LE. The palatopharyngoplasty operation for snoring and sleep apnea: An interim report. Otolaryngol Head Neck Surg 1984; 92: 375–380.
11. Johns MW. A new method for measuring daytime sleepiness: The Epworth sleepiness scale. Sleep 1991; 6: 540–545.
12. Parkes JD, Chen SY, Clift SJ, Dahlitz MT, Dunn G. The clinical diagnosis of the narcoleptic syndrome. J Sleep Res 1998; 7: 41–52.
13. Stradling JR, Davies RJO. Is more NCPAP better? Sleep 2000; 23 (Suppl 4): S150–S153.
14. Shumaker SD, Berzon R. The International Assessment of Health-Related Quality-of-Life: Theory, Translation, Measurement and Analysis. Oxford, UK: Rapid Communication, 1995.
15. Johns MW. Sleepiness in different situations measured by the Epworth sleepiness scale. Sleep 1994; 17: 703–710.
16. Bliwise DL, Carskadon MA, Seidel WF, Nekich JC, Dement WC. MSLT-defined sleepiness and neuropsychological test performance do not correlate in the elderly. Neurobiol Aging 1991; 12: 463–468.
17. Folkerts M, Rosenthal L, Roehrs T, et al. The reliability of the diagnostic features in patients with narcolepsy. Biol Psychiat 1996; 40: 208–214.
18. Zwyghuizen-Doorenbos A, Roehrs T, Schaeffer M, Roth T. Test–retest reliability of the MSLT. Sleep 1988; 11: 562–565.
19. Seidel WF, Dement WC. The multiple sleep latency test: Test–retest reliability. Sleep Res 1981; 10: 284.
20. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16: 297–334.
21. Johns MW. Daytime sleepiness, snoring and obstructive sleep apnea. The Epworth sleepiness scale. Chest 1993; 103: 30–36.
22. Kingshott R, Douglas N, Deary I. Mokken scaling of the Epworth sleepiness scale items in patients with the sleep apnea/hypopnea syndrome. J Sleep Res 1998; 7: 293–294.
23. Liang MG, Fossel AH, Larson ME. Comparisons of five health status instruments for orthopaedic evaluation. Med Care 1990; 28: 632–642.

Address for correspondence: Pa-Chun Wang, Department of Otolaryngology, Cathay General Hospital, 280, Sec. 4, Jen-Ai Rd., 106, Taipei, Taiwan, Republic of China
Phone: +886-2-27082121; Fax: +886-2-27074949
E-mail: drtony@tpts4.seed.net.tw