

The Epworth Sleepiness Scale: Self-administration versus administration by the physician, and validation of a French version

Marta Kaminska MD, Vincent Jobin MD, Pierre Mayer MD, Robert Amyot MD,
Mélanie Perraton-Brillon MD, François Bellemare PhD

M Kaminska, V Jobin, P Mayer, R Amyot, M Perraton-Brillon, F Bellemare. The Epworth Sleepiness Scale: Self-administration versus administration by the physician, and validation of a French version. *Can Respir J* 2010;17(2):e27-e34.

BACKGROUND/OBJECTIVES: The Epworth Sleepiness Scale (ESS) measures sleepiness and is used for, among others, patients with obstructive sleep apnea (OSA). The questionnaire is usually self-administered, but may be physician administered. The aim was to compare the two methods of administration and to validate a French version.

METHODS: Consecutive patients presenting to the sleep clinic at a tertiary care centre completed a self-administered questionnaire containing the ESS. During the medical interview the same day, one of three pulmonologists who specialized in sleep medicine administered the ESS. Correlations with the apnea-hypopnea index and mean sleep latencies were used to assess construct validity, while results of previous self-administered ESS questionnaires in untreated and recently treated OSA patients were used to test reproducibility and longitudinal construct validity, respectively.

RESULTS: In OSA patients, the ESS weakly correlated with the apnea-hypopnea index ($r=0.224$; $P=0.05$; $n=120$) and negatively with mean sleep latency. For untreated patients (test-retest), the mean (\pm SD) average score was unchanged (10.3 ± 6.0 to 10.8 ± 6.5 ; $P=0.35$; $n=56$) after a median of seven months. With continuous positive airway pressure use, the mean score decreased from 12.4 ± 6.8 to 7.6 ± 5.0 after 40.2 months ($P<0.0001$; $n=68$). For all subjects, the ESS score obtained by the physician was less than that of the self-administered result (9.4 ± 5.9 versus 8.5 ± 5.8 ; $P<0.0001$ [paired t test]; $n=188$).

CONCLUSIONS: In a sleep clinic population, the French version of the ESS performed similarly to the English version. However, the systematic underscoring during physician administration may be important to consider in the research setting if questionnaire administration methods are not consistent.

Key Words: Epworth Sleepiness Scale; French; Obstructive sleep apnea; Validation

The Epworth Sleepiness Scale (ESS) was proposed by Johns (1) in 1991 and is widely used in clinical and research settings. It is often used for patients with obstructive sleep apnea (OSA) (2), but also for insomnia (3), and hypersomnia resulting from a variety of etiologies (4-6) including narcolepsy (5) as well as other neurological disorders (6). It was developed as a tool to differentiate patients with excessive daytime sleepiness (EDS) from alert individuals. It consists of a questionnaire containing eight specific situations for which the respondent evaluates their propensity to fall asleep on a scale of 0 to 3. Several translations have been validated (7-11); however, validations of a French language version have yet to be performed.

L'échelle de somnolence d'Epworth : L'autoadministration par rapport à l'administration par le médecin, et la validation d'une version française

HISTORIQUE ET OBJECTIFS : L'échelle de somnolence d'Epworth (ÉSE) mesure la somnolence et est utilisée, entre autres, auprès de patients présentant une apnée obstructive du sommeil (AOS). D'ordinaire, le questionnaire est autoadministré, mais il peut également l'être par le médecin. Les chercheurs souhaitaient comparer les deux modes d'administration et à valider une version française.

MÉTHODOLOGIE : Des patients consécutifs consultant à la clinique du sommeil d'un centre de soins tertiaires ont rempli un questionnaire autoadministré contenant l'ÉSE. Pendant l'entrevue médicale, le même jour, l'un des trois pneumologues spécialisés en médecine du sommeil a également administré l'ÉSE. Les chercheurs ont utilisé les corrélations avec l'indice d'apnée-hypopnée et les latences de sommeil moyennes pour évaluer la validité conceptuelle, tandis qu'ils se sont servi des résultats des questionnaires autoadministrés de l'ÉSE chez les patients présentant une AOS non traitée ou traitée récemment pour vérifier la reproductibilité et la validité conceptuelle longitudinale, respectivement.

RÉSULTATS : Chez les patients ayant une AOS, l'ÉSE était faiblement corrélée avec l'indice d'apnée-hypopnée ($r=0,224$; $P=0,05$; $n=120$) et corrélée négativement avec la latence de sommeil moyenne. Chez les patients non traités (test-retest), l'indice moyen (\pm ÉT) demeurait inchangé ($10,3\pm 6,0$ à $10,8\pm 6,5$; $P=0,35$; $n=56$) après une période médiane de sept mois. En ajoutant une pression positive continue, l'indice moyen fléchissait de $12,4\pm 6,8$ à $7,6\pm 5,0$ au bout de 40,2 mois ($P<0,0001$; $n=68$). Chez tous les sujets, l'indice d'ÉSE obtenu par le médecin était inférieur au résultat du questionnaire autoadministré ($9,4\pm 5,9$ par rapport à $8,5\pm 5,8$; $P<0,0001$ [test t jumelé]; $n=188$).

CONCLUSIONS : Au sein de la population d'une clinique de sommeil, la version française de l'ÉSE donnait des résultats similaires à la version anglaise. Cependant, en milieu de recherche, il serait peut-être important de tenir compte du pointage systématiquement inférieur pendant l'administration par le médecin si les modes d'administration des questionnaires ne sont pas constants.

The English version of the ESS has been well-validated in terms of its reproducibility in normal subjects (1) and in sleep clinic patients (12). It also appears to appropriately reflect change in sleepiness following treatment of OSA (13-15). However, a consistent correlation with a more objective measurement of daytime sleepiness – the multiple sleep latency (MSL) test – was not found or, at best, has been moderate (16-19). The ESS, however, continues to be widely used and its importance is reflected by the fact that the initial publication is the most often cited paper in the sleep literature.

Although the questionnaire was originally developed to be a self-administered tool (1), it is not uncommon for clinicians

Laboratoire du sommeil, Centre Hospitalier de l'Université de Montréal – Hôtel-Dieu, Montréal, Québec

Correspondence: Dr Vincent Jobin, Laboratoire du sommeil, Centre Hospitalier de l'Université de Montréal – Hôtel-Dieu, 3840 St-Urbain, porte 5-243, Montréal, Québec H2W 1T8. Telephone 514-890-8000 ext 15638, fax 514-412-7178, e-mail vincentjobin@hotmail.com

to administer it during the medical interview. Moreover, in the research setting, the ESS may be embedded in a longer questionnaire, with responses obtained from the research subject via an interviewer. These methods of administration have not been validated, and it is unclear whether the score is equivalent to that obtained in the traditional way. In the present prospective study, we validated a French version of the ESS in a sleep clinic population of a tertiary care centre. In addition, we compared scores of the self-administered version with scores obtained via physicians specializing in sleep medicine.

METHODS

Subjects

Two separate groups of subjects were included. All subjects were adult patients of a sleep clinic in a tertiary care referral centre. Group 1 subjects were blind to the study protocol and included new referrals and returning patients. Group 2 subjects were participating in a different study for which they had given consent and included treated OSA patients referred for a maintenance of wakefulness test (MWT) (see the Objective sleepiness measurements section). The only exclusion criterion was the inability to adequately read and understand French to complete the questionnaire.

Procedure

A French version of the ESS questionnaire was given to each patient registering for an appointment with one of three sleep specialists between November 2006 and December 2007. The subjects answered the questionnaire in the waiting room. Thereafter, during the medical encounter, the physician completed a second ESS questionnaire by interviewing the patient without knowledge of the results of the self-administered questionnaire. The three clinicians involved were pulmonologists who specialized in sleep medicine. The study was approved by the Ethics Review Board of the Centre Hospitalier de l'Université de Montréal – Hôtel-Dieu, Montréal, Quebec.

Questionnaire

The ESS was part of a longer questionnaire being used in the sleep clinic at the study hospital. The French version of the ESS was developed by translating the English version textually, and adapting the instructions to make it more easily comprehensible (See Appendix).

A subgroup of subjects had scores from a previously self-administered ESS questionnaire available in their medical chart, from a first visit to the clinic or completed at the time of a sleep study. These previous ESS scores were compared with the score from the current questionnaire, either for a test-retest comparison for OSA subjects who had not been treated or for a pre- versus post-treatment comparison of OSA subjects who had been treated.

Polysomnography

Patients underwent polysomnography (PSG), when indicated, as part of their medical evaluation. Either a full in-laboratory PSG test or an ambulatory simplified test was performed. In-laboratory PSG testing involved an electroencephalography (EEG) recording of (C4/A1, C3/A2, O1/A2, FZ/A1), an electrooculogram (EOG) (R-EOG/A1, L-EOG/A1), chin muscle electromyography and electromyography of the anterior tibialis muscle, and electrocardiogram. Snoring was detected with a

microphone placed at the suprasternal notch. A body position sensor (Braebon Medical Corporation, Canada) attached to a thoracic belt was used to monitor body position. Oxygen saturation was measured by pulse oximetry (OxiMax, Nellcor Puritan Bennett [Melville] Ltd, Canada). Tidal airflow was monitored with a nasal or oronasal pressure cannula (Braebon Medical Corp, Canada). Respiratory efforts were measured via piezoelectric belts placed around the thorax and abdomen. Either a diagnostic 8 h study or a 'split-night' protocol with at least 3 h of sleep recorded in the diagnostic portion was performed. Sandman 7.01 software (Nellcor Puritan Bennett [Melville] Ltd, Canada) was used to analyze data. Two types of ambulatory (American Academy of Sleep Medicine type 3 devices) recording systems were used. The first type of device was the Stardust II device (Respironics, USA), the second was the Suzanne Portable Recording System (Nellcor Puritan Bennett [Melville] Ltd, Canada). Both devices measure nasal airflow with a nasal pressure cannula, oxygen saturation and pulse rate with a finger probe, and respiratory effort via piezoelectric sensors attached to a belt positioned on the chest wall. Data collected with the Stardust device were analyzed using the Stardust host software, and data collected with the Suzanne device were analyzed with the Sandman 7.01 software. Scoring of respiratory events was performed manually by experienced technicians in accordance with accepted criteria (20). An apnea was defined as cessation of airflow for at least 10 s, and a hypopnea as either a clear decrease in airflow of greater than 50% from baseline or of less than 50% but with oxygen desaturation of greater than 3% or an arousal. For PSG studies, arousals were defined using standard EEG-based criteria (20). For portable monitoring, autonomic arousals were defined as a pulse rate increase of at least 5 beats/min. Respiratory effort-related arousals (ie, events not meeting the above criteria but with airflow limitation and an arousal) were scored for clinical decision-making only and were excluded from the event indexes reported here.

Objective sleepiness measurements

Construct validity was assessed in group 2. Subjects completed the questionnaire containing the ESS and underwent an Oxford Sleep Resistance (OSLER) test. Subjects were instructed to press a button on a hand-held box (Stowood Scientific Instruments, England) each time a light signal appeared on a device placed directly in front of them. The light signal was 1 s in duration and repeated every 3 s. A failed test (ie, sleep onset) was defined as the time from the beginning of the test to the point in which seven consecutive signals were missed (MSL_{OSLER}) (21). The number of missed signals per minute was also counted as a measure of impaired vigilance. An EEG was recorded simultaneously to assess EEG-derived MSL (MSL_{EEG}), which was defined according to the American Academy of Sleep Medicine guidelines criteria (22) used in the MWT (ie, one epoch of any stage of sleep). There were four 40 min observation periods. The MSL_{EEGs} reported for MSLs and MSL_{OSLER} were the average of the four trials.

Analysis

Internal consistency was addressed with Cronbach's alpha. Comparisons between groups were performed using the Student's *t* test. Correlations were described with the Pearson correlation coefficient. For the correlation between the apnea-hypopnea

TABLE 1
Demographic and clinical characteristics

	Group 1 subjects				Group 2 subjects
	188	OSA Subjects with previous ESS scores available		68	
		Untreated	Treated		
n	188	56	68	27	
Age, years	54.2±11.9	54.9±11.9	55.7±10.3	51.5±9.0	
Sex, n (male/female)	112/76	46/22	33/23	25/2	
BMI, kg/m ²	31.8±6.9	31.7±5.9	33.2±7.3	34.7±8.7	
Apnea-hypopnea index, per hour	27.3±24.2	24.3±22.4	34.5±23.7	-	
Time between repeat questionnaires*, months	-	7.0 (4.9)	-	-	
Duration of treatment between questionnaires*, months	-	-	40.2 (29.7)	-	

Data presented as mean ± SD unless indicated otherwise. *Data presented as median (interquartile range). BMI Body mass index; ESS Epworth Sleepiness Scale; OSA Obstructive sleep apnea

index (AHI) and the ESS, the pretreatment ESS was always used. The intraclass correlation coefficient (ICC) was used to measure the test-retest reliability (one-way random effects) and the correlation between patient and physician scores (two-way mixed effects).

Comparison of the ESS and MSL was performed by dichotomizing results based on previously published data. The mean (± SD) MSL in normal subjects for the MWT was 35.2±7.8 min (22). Two different thresholds for excessive sleepiness based on the above published normal values were evaluated: the mean minus one SD, so that a 'positive' test for sleepiness was defined as an MSL of 27 min or less, and a 'negative' or normal test as an MSL of greater than 27 min; and the mean minus two SDs, in which a test was considered to be 'positive' with an MSL of 20 min or less, and 'negative' otherwise.

Stata V.8 (Stata, USA) statistical software was used for *t* tests. The software 'R' (23) was used for all other analyses. Correction for multiple testing was performed for correlations and comparisons of subject-physician ESS administration using the false discovery rate procedure as described by Benjamini and Hochberg (24).

RESULTS

Subjects characteristics

A total of 188 subjects (group 1) participated in the present study. In addition, 27 subjects underwent objective sleepiness measurements (group 2). Table 1 summarizes the demographic and clinical characteristics of all study participants. The most common diagnoses in this group of subjects were OSA (n=145), insomnia (n=9), phase delay (n=5), simple snoring (n=3), fibromyalgia (n=3), narcolepsy (n=3, including one who also had OSA) and other/unknown (n=21). The average AHI for OSA subjects in group 1 was 29.3±23.6/h. The average AHI for the 77 OSA subjects who underwent ambulatory testing was 22.2±18.1/h and was 33.4±28.6/h for the 63 subjects who underwent in-laboratory PSG testing. The diagnostic test data for the remaining subjects were unavailable; these

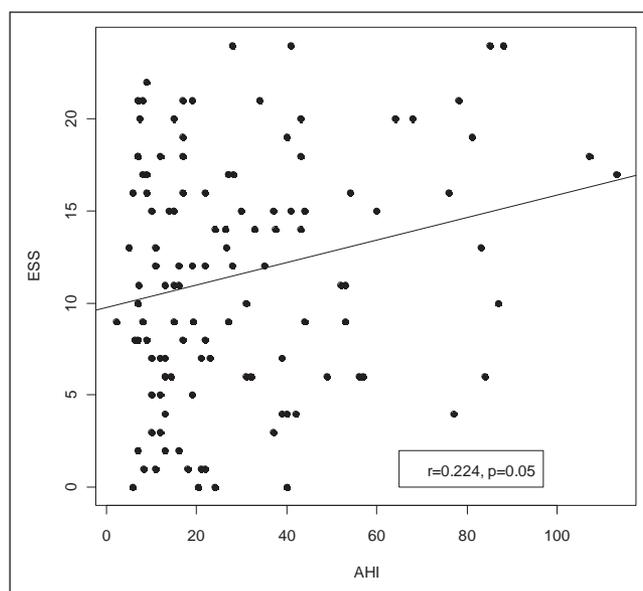


Figure 1 Correlation between the Epworth sleepiness scale (ESS), before any treatment initiation, and the apnea-hypopnea index (AHI) in 124 obstructive sleep apnea subjects

subjects were excluded from any analysis involving the AHI. All group 2 subjects had a diagnosis of OSA established by a treating physician based on diagnostic testing, often performed in a different hospital. Twenty patients were treated with continuous positive airway pressure (CPAP) or bilevel pressure, whereas seven patients refused treatment or were not compliant. They were referred to the Laboratoire du sommeil, Centre Hospitalier de l'Université de Montréal for a MWT.

Internal consistency

A Cronbach's alpha statistic of 0.88 was obtained when the internal consistency of the French version of the ESS was assessed, which signified good coherence without undue redundancy among the different questions.

Construct validity assessment

To assess construct validity, ESS scores were compared with other measures that were believed to be related to sleepiness.

Relationship with OSA severity: Of the 124 OSA subjects, 120 had available data from a diagnostic sleep study as well as an ESS before any treatment for OSA. Among these subjects, there was a weak but significant correlation between the AHI and the ESS ($r=0.224$, $P=0.05$) (Figure 1). The correlation was predominantly observed for those who underwent in-laboratory PSG testing ($r=0.325$, $P=0.05$), whereas it was not significant for those who underwent ambulatory PSG testing ($r=0.072$, $P=NS$). Because the average AHI was lower in those who underwent ambulatory PSG tests, subjects were divided into subgroups to assess the effect of OSA severity on any correlation between the AHI and the ESS: an AHI of 30 or greater and an AHI of less than 30 (regardless of whether ambulatory or in-laboratory PSG testing had been performed). Forty-four subjects had an AHI of 30 or greater, with 27 of these subjects having undergone an in-laboratory PSG test. In this group of severe OSA subjects, there was a correlation that approached

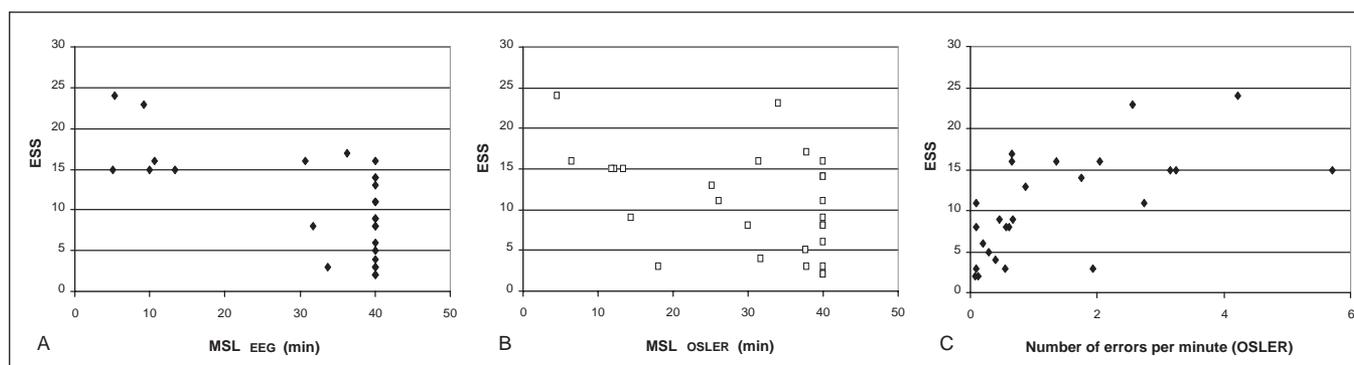


Figure 2) Distribution of Epworth Sleepiness Scale (ESS) scores versus (A) mean electroencephalogram (EEG)-derived sleep latency (MSL_{EEG}) (sleep onset defined as the first epoch of any sleep stage), (B) mean sleep latency on the Oxford Sleep Resistance Test (MSL_{OSLER}) and (C) number of errors per minute on the OSLER, in 27 treated obstructive sleep apnea subjects. A ceiling effect is observed at 40 min, the scheduled end of the test, in panels A and B

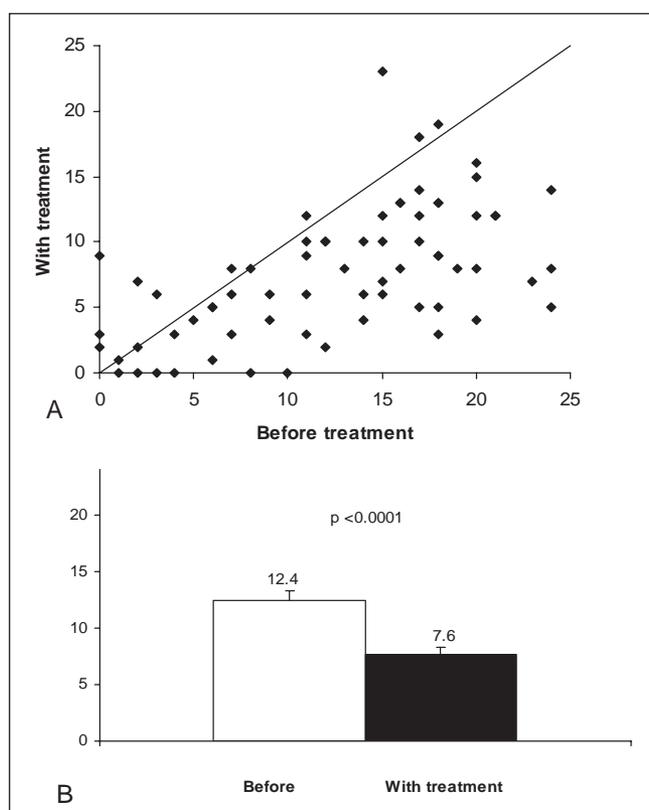


Figure 3) Epworth Sleepiness Scale scores before and after treatment initiation (by continuous positive pressure airway) in a group of 68 obstructive sleep apnea subjects. **A** Individual subjects' scores. Note that most subjects had a decrease in Epworth Sleepiness Scale score with treatment (points to the right of the line). **B** Group averages before and after treatment initiation, demonstrating a decrease in score with treatment (error bars represent standard error of the mean)

significance between the AHI and the ESS ($r=0.311$; $P=0.07$). There was no correlation in the group of 76 subjects with lower AHIs ($r=0.044$, $P=NS$).

Relationship with objective measures of sleepiness: The ESS was compared with the MSL in a group of 27 treated OSA subjects. Negative correlations were found between the ESS and both the MSL_{EEG} ($r=-0.665$, $P=0.02$, Figure 2A) as well as

the MSL_{OSLER} ($r=-0.429$, $P=0.05$, Figure 2B). In addition, the number of errors per minute on the OSLER was positively correlated with the ESS ($r=0.630$, $P=0.02$, Figure 2C).

Because correlation may not be the optimal statistical method of choice when a ceiling effect exists, the sensitivity and specificity of the ESS was calculated. An ESS score of more than 10 was considered to be 'positive' and was compared with the MSL_{EEG} and the MSL_{OSLER} . Two cut-off point MSLs were tested: 27 min and 20 min (see Methods). Using a MSL_{EEG} of less than 27 min as 'positive', the ESS was 100% sensitive (six of six) and 62% specific (13 of 21) at detecting excessive sleepiness. The positive predictive value (PPV) and negative predictive value (NPV) were 43% (six of 14) and 100% (13 of 13), respectively. For a MSL_{EEG} of less than 20 min, the same values were obtained. For an MSL_{OSLER} of less than 27 min, the ESS was 78% sensitive (seven of nine) and 61% specific (11 of 18) at detecting abnormal sleepiness, with a PPV of 50% (seven of 14) and an NPV of 85% (11 of 13). Using a 'positive' MSL_{OSLER} of less than 20 min, the ESS sensitivity was 71% (five of seven) and the specificity 55% (11 of 20), with a PPV of 36% (five of 14) and an NPV of 85% (11 of 13).

Response to treatment: Longitudinal construct validity was assessed in a group of 68 treated OSA subjects who had an ESS score available before treatment initiation. The ESS improved in the majority of subjects after a median duration of treatment of 40.2 months (Figure 3A). The mean ESS was 12.4 ± 6.8 before treatment and 7.6 ± 5.0 with treatment ($P < 0.0001$) (Figure 3B). Compliance data from CPAP machines' memory cards were available for 27 subjects, of whom 21 (78%) were using their CPAP device at least 4 h each night. Among the compliant subjects, the mean ESS decreased from 13.1 ± 6.5 to 7.6 ± 4.1 ($P=0.0007$).

Reproducibility: Reproducibility was assessed in a group of OSA subjects who were not being treated at the time of the study, mostly because they were waiting to participate in a CPAP titration study, with some having declined treatment. Fifty-six subjects had a previous ESS in the chart. The median time between the two questionnaires was 7.0 months, with presumed clinical stability because no treatment had been initiated. Scores were similar on both iterations of the test for most subjects (Figure 4A), with a strong ICC between scores (0.847). The group means were equivalent (10.3 ± 6.0 versus 10.8 ± 6.5 ; $P=0.35$) (Figure 4B). The absolute difference in score between the two tests was

greater than 4 in 27% of subjects. Four subjects who initially had a score of more than 10 had a score of 10 or less on the second iteration; six subjects who initially had a score of 10 or less subsequently had a score of more than 10. Thus, overall, 10 of 56 (18%) subjects changed status, based on the commonly accepted threshold of greater than 10 as being 'abnormally sleepy'.

Methods of administration

The effect of the ESS administration method was assessed by comparing scores from the self-administered questionnaire with scores obtained by a physician specializing in sleep medicine during a medical interview. Three physicians participated. The number of subjects seen by each physician was as follows: physician 1 (n=61), physician 2 (n=21) and physician 3 (n=106). On average, scores from the self-administered questionnaire were higher than those obtained by the physician (9.4 ± 5.8 versus 8.3 ± 5.8 ; $P < 0.0001$) (Figures 5A and 5B); there was, however, a strong correlation between the two (ICC = 0.835). A similar subject-physician difference was present for all three physicians (although it failed to reach statistical significance for physician 1) (Figure 5C). This difference was also found regardless of diagnosis, that is, for OSA and non-OSA subjects alike (data not shown).

Each question was examined separately to determine whether any one question was predominantly responsible for most of the difference between the two scores. The differences ranged from 0.085 ± 0.80 (question 3 – "in a public place") to 0.26 ± 1.11 (question 1 – "reading").

DISCUSSION

The present study validated a French version of the ESS. In addition, it showed that when the ESS was administered to the patient by a physician in the context of a medical interview, the resulting score was somewhat lower than the score obtained by the self-administered questionnaire.

Questionnaire validation

The ESS is intended to measure only one cohesive attribute, sleep propensity. In the original validation paper, Cronbach's alpha – used to measure internal consistency – was 0.88 in OSA subjects. The same alpha value was obtained in the current study in a mixed group of subjects from a sleep clinic using a French version of the ESS, suggesting a high degree of internal consistency and little redundancy within the questionnaire.

An instrument is considered to have a high degree of construct validity when it demonstrates strong correlations with other measurements related to the same attribute. In OSA, sleepiness should correlate with disease severity, for example, as measured by the AHI. In the present study, we found a weak correlation between the ESS score and the AHI in OSA subjects. A correlation was primarily found in those with severe OSA (ie, AHI of 30 or greater) and not in those with mild or moderate OSA. One of the limitations of the present study was that diagnostic testing for OSA was not uniform, with some subjects having undergone in-laboratory PSG testing while others had simplified ambulatory testing, which may have weakened the correlation between the ESS and AHI. Portable monitoring tends to be least accurate at lower AHI values

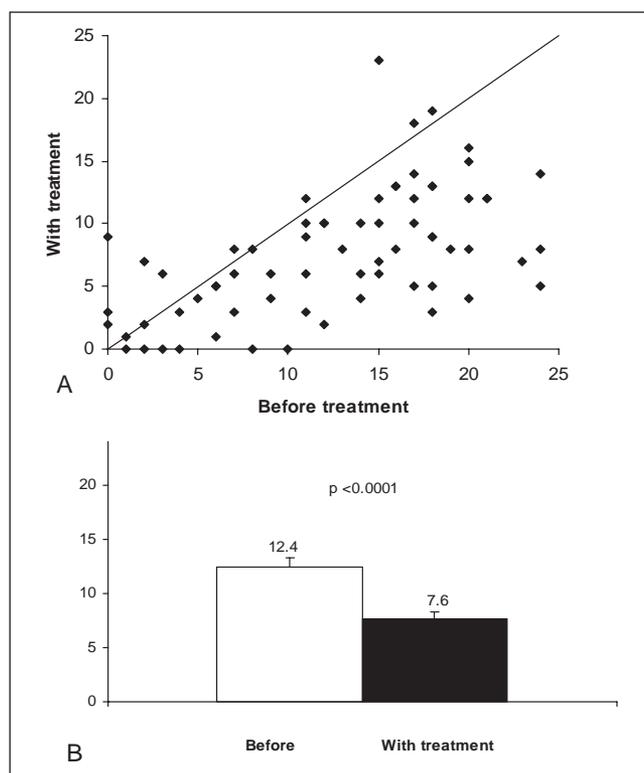


Figure 4 The Epworth sleepiness scale (ESS) test-retest in a group of 56 untreated obstructive sleep apnea subjects. **A** Individual subjects' scores. **B** Mean scores on the two iterations, with a median of 7.0 months in between (error bar represents standard error of the mean)

(25), which may be reflected in our finding, because the greatest scatter in AHI versus ESS score was at the lowest range of AHI values. The higher AHI in subjects having undergone in-laboratory testing may reflect an underestimation of the index by ambulatory testing either because a different denominator was used in calculating the AHI (total recording versus total sleep time), or due to the inability of the portable monitor to detect EEG arousals. However, autonomic arousals were scored in the simplified studies, and although they may not be directly comparable with EEG-based arousals, they appear to have reasonable accuracy in similar circumstances (26,27). An alternative explanation for the higher AHI in subjects who underwent in-laboratory testing is that subjects with a higher clinical probability of severe OSA, or with a greater clinical urgency to initiate treatment, may have been preferentially referred to a 'split-night' protocol and, therefore, may truly represent a more severe OSA group. Our findings are consistent with several previous studies (28,29) attempting to correlate the original English version of the ESS with the AHI. In the Sleep Heart Health Study (28), ESS scores increased with the number of apneas and hypopneas per hour, but there was little difference in the proportion of subjects with an ESS score of greater than 10 between those with a respiratory disturbance index of less than 5/h and those with a respiratory disturbance index of 30/h or greater (21% and 35%, respectively). Indeed, aside from the AHI, sleepiness in OSA is known to be related to several clinical features such as the presence of respiratory disease, self-reported sleep duration (29), depression and lack of regular exercise (30). In addition, some apneic subjects appear not to

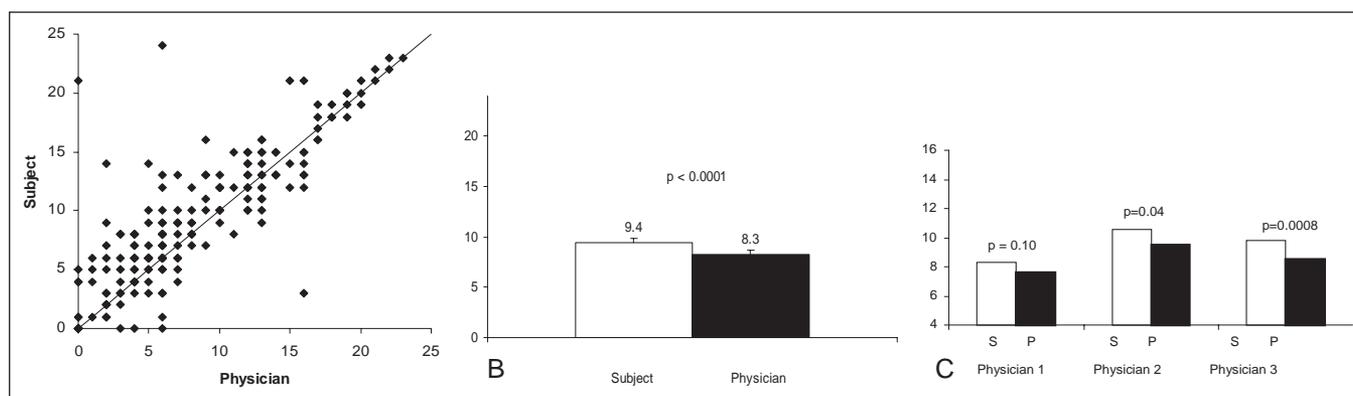


Figure 5 Epworth Sleepiness Scale scores from the self-administered questionnaire ('subject') versus scores obtained by the physician during a medical interview ('physician'). **A** Scores for the 188 individual subjects. **B** Average scores for the two modes of administration (error bars represent standard error of the mean). **C** Average Epworth Sleepiness Scale scores according to method of administration for three different physicians

recognize their sleepiness because it develops inconspicuously and is present chronically. A response shift regarding sleepiness in OSA has been demonstrated after CPAP treatment (31).

We found that the ESS demonstrated relatively good sensitivity but poor specificity in detecting objectively measured EDS. Sensitivity and specificity of the ESS were best for the MSL_{EEG} rather than the MSL_{OSler} . Moreover, none of the subjects with an ESS score of less than 10 demonstrated any objective pathological sleepiness, as defined by the thresholds mentioned earlier; there were no 'false negatives'. Thus, a normal ESS score may obviate the need for objective assessment of sleepiness in treated OSA subjects, with the understanding that extrapolation to other populations is unjustified. Conversely, a high ESS score did not predict excessive objective sleepiness (ie, a short MSL). Significant but weak correlations were found between the ESS and objective sleepiness measures (MSL_{OSler} , MSL_{EEG} and number of errors per minute on the OSLER). These results are consistent with previous studies using the English version of the questionnaire (16-19,32). Objective sleepiness tests do not appear to measure exactly the same phenomena as the ESS. Whether the MWT and the MSL test should be the gold standards for measuring EDS has been debated (33).

The simultaneous measurement of MSL using MSL_{OSler} and MSL_{EEG} in the present study may be considered to be a limitation. The OSLER could theoretically have created a distraction that delayed sleep onset as measured on the MSL_{EEG} . However, this is unlikely because the OSLER provides a repetitive, monotonous signal with minimal stimulation, and the MSL is known to be similar, whether the MWT and OSLER tests are performed simultaneously (34) or separately (21).

We have found the French version of the ESS to be reflective of the effects of treatment with CPAP in OSA subjects. Indeed, after a median of 40.2 months of treatment, most subjects showed an improvement in the ESS, with the mean decreasing from 12.4 to 7.6, confirming longitudinal construct validity of the questionnaire. We had compliance data only on a subset of patients, who showed significant response to treatment as well. Results for the entire group can be considered to represent an intention to treat analysis, although our population was drawn from a sleep clinic and, therefore,

only subjects who came for follow-up were included. Thus, the effect found represents the average decrease in ESS score in a typical sleep clinic population in a referral centre. Our results are comparable with those found for the English version of the ESS. Changes with CPAP treatment in OSA subjects in different studies were from 15 ± 6 to 7 ± 5 (14), from 13 ± 3 to 8 ± 4 (15), and from a mean (\pm standard error) of 12.1 ± 0.6 to 5.6 ± 0.5 (13).

Reproducibility was initially demonstrated for the English version in a group of healthy medical students (1). In the current study, a test-retest approach in untreated OSA subjects was used to examine the reproducibility of the French version of the ESS in this patient population. Despite good reproducibility overall, we found a difference of greater than 4 in 27% of subjects on retesting. Nguyen et al (12) found this same difference in 23% of their subjects, who were, similar to ours, from a sleep clinic population (12). A limitation of our study was that the two iterations of the questionnaire were not studied prospectively. Stability of OSA between the two iterations was assumed, because no treatment had been initiated, but could not be verified. Other potential confounders such as change in body mass index, sleep duration or medication, could not be accounted for in the present study, and may have caused a true difference in the sleepiness of some subjects. However, despite an unchanged clinical situation, it cannot be excluded that the questionnaire lacks reproducibility in some subjects. The common approach of using a cut-off of greater than 10 to classify someone as 'abnormal' versus 'normal' should be used with caution because patients may change status on different iterations of the self-administered questionnaire, as we found in 18% of our subjects. Nevertheless, there was no evidence of a systematic increase or decrease in score over time in the group, thereby providing supportive evidence for the reproducibility of the French version of the ESS questionnaire.

Administration by the physician

We found that when the questionnaire was administered by a physician, the score was lower than that from a self-administered questionnaire. None of the eight questions seemed spared from the effect, although the greatest difference was for the first question ("sitting and reading"). It is not clear whether this was

due to the question itself, whether it was the first question or due to chance. There are several possible explanations for the observed difference between the self- and physician-administered questionnaire score. The self-administered questionnaire is answered in private, with some anonymity. Subjects may feel freer to expand on their perceived problems under these circumstances. On the other hand, in the setting of an interview, a social desirability bias may have been introduced. Subjects are likely to under-report behaviours that they believe are socially less acceptable. This is apparently true even when the interviewer is the treating physician. Kumru et al (35) found that partners tended to score subjects' sleepiness higher than the subjects themselves on the ESS, with an average score difference of 1.2. This may have been due to a difference in perception between the subject and the partner or, again, due to a social desirability bias on the part of the subjects.

The delay between the self-administered questionnaire and the physician interview in the present study was short, which may have led to recall bias. The effect of such a bias would have made the scores more similar to each other. The finding of a significant difference between them is, therefore, all the more noteworthy. This difference was similar for all three sleep physicians participating in the current study and, therefore, was not due to an individual misperception of patients' responses. However, the present study was not randomized, and the self-administered questionnaire was always completed first, which may have influenced the answers to the physician-administered questionnaire. This difference, although consistent, is on average small (mean score difference of 1.1) and unlikely to be relevant in the clinical context. Moreover, the difference between ESS scores before and after treatment was much greater, such that the method of ESS administration for the purpose of clinical follow-up may not be crucial. However, in the context of research, in which small differences can be statistically significant, the ESS must be administered in a consistent fashion, with the method clearly specified so as to avoid biased results.

REFERENCES

- Johns MW. A new method for measuring daytime sleepiness: The Epworth sleepiness scale. *Sleep* 1991;14:540-5.
- Marshall NS, Barnes M, Travier N, et al. Continuous positive airway pressure reduces daytime sleepiness in mild to moderate obstructive sleep apnoea: A meta-analysis. *Thorax* 2006;61:430-4.
- Roth T, Drake C. Evolution of insomnia: Current status and future direction. *Sleep Med* 2004;5(Suppl 1):S23-S30.
- Dauvilliers Y. Differential diagnosis in hypersomnia. *Curr Neurol Neurosci Rep* 2006;6:156-62.
- Fry JM. Treatment modalities for narcolepsy. *Neurology* 1998;502(Suppl 1):S43-S48.
- Thorpy MJ. Sleep disorders in Parkinson's disease. *Clin Cornerstone* 2004;6(Suppl 1):A:S7-15.
- Bloch KE, Schoch OD, Zhang JN, Russi EW. German version of the Epworth Sleepiness Scale. *Respiration* 1999;66:440-7.
- Chen NH, Johns MW, Li HY, et al. Validation of a Chinese version of the Epworth sleepiness scale. *Qual Life Res* 2002;11:817-21.
- Chiner E, Arriero JM, Signes-Costa J, Marco J, Fuentes I. [Validation of the Spanish version of the Epworth Sleepiness Scale in patients with a sleep apnea syndrome]. *Arch Bronconeumol* 1999;35:422-7.
- 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;12:161-8.
- Tsara V, Serasli E, Amfilochiou A, Constantinidis T, Christaki P. Greek version of the Epworth Sleepiness Scale. *Sleep Breath* 2004;8:91-5.
- Nguyen AT, Baltzan MA, Small D, Wolkove N, Guillon S, Palayew M. Clinical reproducibility of the Epworth Sleepiness Scale. *J Clin Sleep Med* 2006;2:170-4.
- Ballester E, Badia JR, Hernandez L, et al. Evidence of the effectiveness of continuous positive airway pressure in the treatment of sleep apnea/hypopnea syndrome. *Am J Respir Crit Care Med* 1999;159:495-501.
- Engleman HM, sgari-Jirhandeh N, McLeod AL, Ramsay CF, Deary IJ, Douglas NJ. Self-reported use of CPAP and benefits of CPAP therapy: A patient survey. *Chest* 1996;109:1470-6.
- Engleman HM, Kingshott RN, Wraith PK, Mackay TW, Deary IJ, Douglas NJ. Randomized placebo-controlled crossover trial of continuous positive airway pressure for mild sleep apnea/hypopnea syndrome. *Am J Respir Crit Care Med* 1999;159:461-7.
- Chervin RD, Aldrich MS. The Epworth Sleepiness Scale may not reflect objective measures of sleepiness or sleep apnea. *Neurology* 1999;52:125-31.
- Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep* 1994;17:703-10.
- Benbadis SR, Mascha E, Perry MC, Wolgamuth BR, Smolley LA, Dinner DS. Association between the Epworth sleepiness scale and the multiple sleep latency test in a clinical population. *Ann Intern Med* 1999;130(4 Pt 1):289-92.
- Chervin RD, Aldrich MS, Pickett R, Guilleminault C. Comparison of the results of the Epworth Sleepiness Scale and the Multiple Sleep Latency Test. *J Psychosom Res* 1997;42:145-55.
- Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. *Sleep* 1999;22:667-89.

SUMMARY

We have shown that, in OSA patients, the French version of the ESS has validity and reliability comparable with those reported for the English version. It has good overall reproducibility and reflects the response to CPAP treatment. Interestingly, in treated OSA subjects, a score of 10 or less appears to exclude objectively measured EDS. Moreover, we found that scores were lower when the ESS was administered by a physician in the context of the medical interview than the scores obtained with the self-administered questionnaire. This creates the need to be consistent and explicit about the method of test administration, particularly when used in research studies in which relatively small differences can create significant bias in results.

ACKNOWLEDGEMENTS: The authors thank Médigaz, OSR Médical and Apnair for their unrestricted grants to the "Fondation du CHUM", as well as Anne-Marie Laurin, Chief Technician, and all staff of the Sleep Laboratory at the CHUM Hôtel-Dieu.

APPENDIX

Échelle de somnolence d'Epworth	
Quel risque avez-vous de vous assoupir ou de vous endormir dans les situations suivantes, sans considérer les moments où vous vous sentez fatigué(e) ? Ces questions s'appliquent à vos habitudes de vie actuelles, dans des conditions habituelles. Si vous n'avez pas effectué récemment certaines activités, essayez d'imaginer comment vous auriez réagi. Utilisez l'échelle ci-dessous en choisissant le chiffre le plus approprié pour chaque situation.	
0 : aucun risque de m'assoupir ou de m'endormir 1 : faible risque de m'assoupir ou de m'endormir 2 : risque modéré de m'assoupir ou de m'endormir 3 : risque élevé de m'assoupir ou de m'endormir	
Situations	Scores (0, 1, 2 ou 3)
Lire en position assise	_____
Regarder la télévision	_____
Être assis(e) inactif(ve) dans un lieu public (théâtre, réunion, par exemple)	_____
Être assis(e) en tant que passager(ère) dans un véhicule pour une période d'une heure sans arrêt	_____
Être étendu(e) l'après-midi lorsque les circonstances le permettent	_____
Être assis(e) en parlant avec quelqu'un	_____
Être assis(e) tranquille après un repas sans boisson alcoolique	_____
Dans une voiture arrêtée quelques minutes à un feu de circulation ou dans la circulation	_____
Total :	_____

21. Bennett LS, Stradling JR, Davies RJ. A behavioural test to assess daytime sleepiness in obstructive sleep apnoea. *J Sleep Res* 1997;6:142-5.
 22. Littner MR, Kushida C, Wise M, et al. Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep* 2005;28:113-21.
 23. R Development Core Team. R: A language and environment for statistical computing. R foundation for statistical computing 2008. <<http://www.R-project.org>> (Accessed on November 6, 2008).
 24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Stat Soc.* 1995;57:289-300.
 25. Yin M, Miyazaki S, Ishikawa K. Evaluation of type 3 portable monitoring in unattended home setting for suspected sleep apnea: Factors that may affect its accuracy. *Otolaryngol Head Neck Surg* 2006;134:204-9.
 26. Adachi H, Mikami A, Kumano-go T, et al. Clinical significance of pulse rate rise during sleep as a screening marker for the assessment of sleep fragmentation in sleep-disordered breathing. *Sleep Med* 2003;4:537-42.
 27. Pitson D, Stradling J. Autonomic markers of arousal during sleep in patients undergoing investigation for obstructive sleep apnoea, their relationship to EEG arousals, respiratory events and subjective sleepiness. *J Sleep Res* 1998;7:53-9.
 28. Gottlieb DJ, Whitney CW, Bonekat WH, et al. Relation of sleepiness to respiratory disturbance index: The Sleep Heart Health Study. *Am J Respir Crit Care Med* 1999;159:502-7.
 29. Kapur VK, Baldwin CM, Resnick HE, Gottlieb DJ, Nieto FJ. Sleepiness in patients with moderate to severe sleep-disordered breathing. *Sleep* 2005;28:472-7.
 30. Basta M, Lin HM, Pejovic S, Sarrigiannidis A, Bixler E, Vgontzas AN. Lack of regular exercise, depression, and degree of apnea are predictors of excessive daytime sleepiness in patients with sleep apnea: Sex differences. *J Clin Sleep Med* 2008;4:19-25.
 31. Chin K, Fukuhara S, Takahashi K, et al. Response shift in perception of sleepiness in obstructive sleep apnea-hypopnea syndrome before and after treatment with nasal CPAP. *Sleep* 2004;27:490-3.
 32. Philip P, Sagaspe P, Taillard J, et al. Maintenance of Wakefulness Test, obstructive sleep apnea syndrome, and driving risk. *Ann Neurol* 2008;64:410-6.
 33. 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.
 34. Krieger AC, Ayappa I, Norman RG, Rapoport DM, Walsleben J. Comparison of the maintenance of wakefulness test (MWT) to a modified behavioral test (OSLER) in the evaluation of daytime sleepiness. *J Sleep Res* 2004;13:407-11.
 35. Kumru H, Santamaria J, Belcher R. Variability in the Epworth Sleepiness Scale score between the patient and the partner. *Sleep Med* 2004;5:369-71.
-



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

