

Reliability and validity of a Norwegian version of the Epworth sleepiness scale

Kornelia K. Beiske · Frank N. Kjelsberg ·
Espen A. Ruud · Knut Stavem

Received: 7 March 2008 / Revised: 9 May 2008 / Accepted: 14 May 2008 / Published online: 17 June 2008
© Springer-Verlag 2008

Abstract

Objective The objective of this study was to assess the reliability and validity of a Norwegian version of the self-administered Epworth sleepiness scale (ESS).

Materials and methods Two samples responded to the ESS: (1) 226 patients previously evaluated for obstructive sleep apnea, of whom 51 also responded to a retest 2 weeks later, and (2) 37 ambulant patients complaining of excessive daytime sleepiness, who were referred to multiple sleep latency testing (MSLT). We assessed internal consistency reliability with Cronbach's alpha and test–retest reliability with weighted kappa (Kw) or an intraclass correlation coefficient (ICC). The validity of the Norwegian ESS was assessed by correlating ESS item and total scores with the number of times a patient fell asleep and the mean latency found on the MSLT.

Results Internal consistency reliability, as assessed with Cronbach's alpha, was 0.84 ($n=154$). Test–retest reliability for the eight ESS items ranged from Kw of 0.61 to 0.80 ($n=50$) and for the total score. ICC was 0.81. There was only fair to moderate correlation of ESS item and total scores with MSLT variables, mainly in a subset of patients with total ESS score >10 .

Conclusions The Norwegian version of the ESS had acceptable internal consistency and test–retest reliability. The association of the ESS items and total score with the MSLT was only fair to moderate, in line with previous studies.

Keywords Obstructive sleep apnea · Epworth sleepiness scale · Multiple Sleep Latency Test · Reliability · Validity · Norwegian

K. K. Beiske (✉)
Department of Neurology, Medical Division,
Akershus University Hospital,
Lørenskog, Norway
e-mail: kornelia.beiske@ahus.no

F. N. Kjelsberg · E. A. Ruud · K. Stavem
Department of Pulmonary Medicine, Medical Division,
Akershus University Hospital,
Lørenskog, Norway

K. Stavem
Helse Sør-Øst Health Services Research Centre,
Akershus University Hospital,
Lørenskog, Norway

K. Stavem
Faculty of Medicine, University of Oslo,
Oslo, Norway

Introduction

The Epworth sleepiness scale (ESS) is a popular, self-administered questionnaire developed for assessing subjective average sleep propensity, which is inexpensive and easy to administer [1–5]. In spite of the widespread use of the ESS, there is limited documentation on the reliability and validity of the questionnaire. Several studies have expressed conflicting opinions about the association between ESS score and multiple sleep latency as measured by the multiple sleep latency test (MSLT) [6]. Some studies show a moderate association between ESS score and mean sleep latency of the MSLT [7], while the majority show no significant [8–10] or a negative association [4, 5] between ESS score and mean sleep latency.

The ESS has been translated to many languages; however, the number of different translated versions of

ESS in use today is unclear. There are only a few articles documenting the test–retest reliability and validity of translated versions of the ESS [2, 11, 12].

The objective of this study was to assess the internal consistency and the test–retest reliability and validity of a Norwegian version of the ESS using two populations: (1) subjects who were previously evaluated for sleep apnea syndrome and (2) subjects who presented with complaints of excessive daytime sleepiness (EDS) and were referred to and successfully completed an ambulant MSLT.

Materials and methods

Study design and subjects

This study is based on data collection from two samples: (1) a cross-sectional postal survey of subjects who were previously evaluated for obstructive sleep apnea (OSA) and (2) consecutive patients with complaints of EDS, who were referred by a general practitioner, a specialist in neurology, or otolaryngology, and successfully completed an ambulant MSLT. This study was approved by The Regional Committee for Medical and Health Research Ethics.

Patients evaluated for obstructive sleep apnea

A questionnaire was mailed to all 242 patients evaluated for continuous positive airway pressure (CPAP) for treatment of OSA, without any known comorbid conditions, who were referred by a general practitioner, a specialist in neurology, or otolaryngology, and admitted to the pulmonary unit of the Akershus University Hospital from 1993 to 1999 [13]. Patients with a discharge diagnosis of the International Classification of Diseases, ninth revision (ICD-9) code 780.5 (1993–1998) or ICD-10 code G47.3 (1999) were eligible for the study.

The questionnaire included the ESS, items about demographics, smoking, and use of CPAP equipment. Pilot testing of the questionnaire in nine patients with OSA resulted in minor adjustments. A reminder was sent after 2 weeks and another after 4 weeks. We reviewed the medical records of the patients, abstracting relevant information and data from a sleep study before the hospital admission [13]. The apnea–hypopnea index (the total number of episodes of apnea and hypopnea divided by the number of hours of sleep) was calculated for each patient, as a measure of disease severity.

Among the respondents in the survey, 160 consented to be contacted again. From these 160 respondents, we randomly selected 90 patients who were asked by telephone to participate in a test–retest assessment of the question-

naire, including the ESS, 2 weeks apart. Sixty-nine patients agreed to participate, of whom 65 responded to the first questionnaire and 51 to the second. The mean time between assessments was 18 days.

Patients referred to sleep laboratory for multiple sleep latency test

All subjects referred to Akershus University Hospital between 2003 and 2005 for ambulant MSLT were eligible for the study. They were referred for EDS by a general practitioner, a specialist in neurology, or otolaryngology. Those eligible for the study were asked to fill out a questionnaire, including the ESS. Forty-six subjects agreed to participate and were given the questionnaire while at the department being fitted and instructed in use of the MSLT equipment. They filled out the questionnaire at home and returned it the next day along with the MSLT equipment. Of the 46 participants 37 executed the MSLT correctly and were subsequently included in the study.

Epworth sleepiness scale

The ESS is a widely used, English-written, self-administered questionnaire developed for assessing subjective average sleep propensity [1, 2]. The questionnaire assesses the level of general sleepiness during eight real life situations where subjects are asked to rate their chance of dozing in recent times in each situation on a scale of 0 to 3. The responses are added together to produce an ESS score ranging from 0 to 24 [1]. It has been reported that, at a cut-off score of >10, the ESS has a sensitivity of 93.5% and a specificity of 100% in distinguishing EDS from normal daytime sleepiness [14].

In this study, a Norwegian version of the ESS was used for assessing daytime sleepiness. The English version of the ESS was translated into Norwegian using a standardized translation procedure. The English version was first translated to Norwegian by two native speaking Norwegian physicians who were fluent in English and then translated back to English by a translator, whose mother tongue was English, for comparison. Written consent, from the copyright holder Murray Johns, was given for the use of this Norwegian version of the ESS.

Overnight polysomnography

Patients who were evaluated for sleep apnea underwent an ambulant sleep apnea recording using the Embla system (Embla; Medcare-Flaga; Reykjavik, Iceland). In these studies, signals from airflow (using thermistors), arterial oxygen saturation, heart rate, abdominal and thoracic

respiratory movements, body position, and snoring were recorded.

These polysomnography (PSG) recordings were scored the next day by doctors specializing in clinical neurophysiology using criteria from the 1999 American Academy of Sleep Medicine Task Force [15].

Multiple sleep latency test

The MSLT is considered to be the gold standard for the diagnosis of EDS [2] and was used to validate this Norwegian version of the ESS.

The participants referred to MSLT also underwent a PSG recording using the ambulant Embla system, including surface electrode recordings of the electroencephalogram (C3, C4, O1, and O2), electro-oculograms (EOG1 and EOG2), electrocardiogram, as well as m. submental and bilateral m. tibialis anterior electromyograms. Before the MSLT and PSG, all patients filled out sleep logs, for a duration of 2 weeks, where information about use of medications was also described. Patients were also required to fill out a sleep log the day and night of the recordings. In conjunction with the MSLT, patients wrote down if they were presently taking any medication, when and how fast they fell asleep, and whether or not they dreamt during the nap periods. In conjunction with the PSG, patients described when they had their last meal of the day, if they drank alcohol or took any medications the night of the recording, when they went to bed, when they turned off the lights, how many hours they felt they slept, if they were up and out of bed at any time during the night, and if they had any problems with the Embla equipment.

The MSLT was completed during the daytime, where the patients were given five nap opportunities, with test durations of 20 min, 2 h apart throughout the day. Patients themselves marked the start and stop time of each nap period by pressing the event button on the Embla equipment. They were required to stay awake between these napping sessions to correctly complete the test. Patients who did not comply, by either sleeping in between the set napping sessions or by not indicating their start and stop napping times with the event button, were not included in this study, as the MSLT was considered to be incorrectly done. The PSG was then carried out the following night. This practice is a variation of the accepted convention [16, 17], but makes these studies more convenient and economical for both the patient and the laboratory. The recordings were analyzed and scored the next day by doctors specializing in the field of clinical neurophysiology, using the Rechtschaffen and Kales standardized scoring system for sleep stages [18]. Three doctors were responsible for the scoring of these tests, but 95% of the recordings included in

this study were scored by only one of these doctors. Due to this fact, inter-scoring reliability is high.

Statistical analysis

Descriptive statistics are presented using means (SD) or number (%). Internal consistency reliability of the ESS total score was assessed using Cronbach's coefficient alpha in the total sample of patients evaluated for OSA. Test-retest reliability of the ESS items was assessed using weighted kappa, with quadratic disagreement weights kappa statistics indicate the proportion agreement between observers after correction for the agreement expected by chance, with 1 indicating full agreement, 0 equal to chance alone, and -1 complete disagreement [19]. Test-retest reliability for the aggregated ESS total score was assessed using an intraclass correlation coefficient.

We assessed the construct validity of ESS by Spearman's rank correlation of ESS item and total scores with variables from the MSLT: number of times the patient fell asleep and mean latency to sleep (in minutes). When discussing the strength of agreement correlations, we used the following nomenclature from agreement statistics [20]: <0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, 0.81–1.00 very good. A moderate to good correlation would support the construct validity of ESS. We also assessed this separately in a subset of patients with pathological ESS score, i.e., ESS > 10.

We further assessed criterion-related validity in the total sample of patients that were evaluated for OSA by investigating whether the ESS items and total score could discriminate between groups of patients selected according to snoring (yes vs. no), apnea-hypopnea index above/below the sample median (>24 per hour vs. ≤24 per hour) and whether the patients were in possession of a CPAP machine (yes vs. no). Similarly, in the MSLT group, we assessed criterion-related validity using self-reported information about snoring (yes vs. no), sleeping during the day (yes vs. no), and sleepiness while driving (yes vs. no) as criteria.

Groups were compared using the *t* test for independent samples. We chose a 5% significance level. The SPSS statistical software version 11.5 (SPSS, Chicago, IL, USA) or Stata version 8.2 (Stata Corp., College Station, TX, USA) were used for the analyses.

Results

Descriptive statistics comparing the two groups of patients are presented in Table 1. Among patients complaining of EDS who underwent MSLT ($n=37$), the mean age was 43 years (age range, 32–54 years), and 32% were women. For

patients who were previously evaluated for OSA ($n=178$), the mean age was 56 years (age range, 45–67 years), and 24% were women.

Reliability of the Norwegian version of the Epworth sleepiness scale

Internal consistency reliability, as assessed with Cronbach's alpha, was 0.84 ($n=154$). Test–retest reliability of the Norwegian version of ESS, as assessed with weighted kappa in a subset of patients evaluated for OSA, ranged from 0.61 to 0.80 for the eight different ESS items ($n=50$) (Table 2). For the aggregate ESS total score, the test–retest reliability was 0.81, as assessed with an intraclass correlation coefficient.

Validity of the Norwegian version of the Epworth sleepiness scale

In the total sample of participants referred for EDS ($n=37$), there was only a fair to moderate correlation between the number of times a patient fell asleep and the ESS items of “sitting quietly after a lunch without alcohol” and “lying down to rest in the afternoon when circumstances permit,”

and poor correlation for the remaining individual ESS items and the ESS items for the total sample (Table 3).

In a subset of patients ($n=25$) with ESS scores > 10 , we found a fair correlation between the number of times a patient fell asleep and the ESS items of “sitting quietly after a lunch without alcohol,” “in a car, while stopped for a few minutes in traffic,” “sitting and reading,” and the ESS total score. There was a moderate correlation between the number of times a patient fell asleep and the ESS item of “watching TV.” The remaining ESS items showed poor correlation to the number of times a patient fell asleep. In this subset, we also found a fair correlation between the mean latency to sleep and the ESS items of “lying down to rest in the afternoon when circumstances permit,” “sitting and talking to someone,” “sitting quietly after a lunch without alcohol,” and total ESS score. The remaining ESS items showed poor correlation with the mean latency to sleep (Table 3).

When comparing ESS items and total scores according to presence or absence of snoring, sleeping during the day, and sleepiness while driving in patients complaining of EDS who underwent MSLT, as well as ESS item and total scores according to snoring status, $AHI < 24 \text{ h}^{-1}$ vs. $AHI > 24 \text{ h}^{-1}$, and possession of a CPAP machine, in patients evaluated for OSA, one finds few significant correlations (Tables 4 and 5).

Table 1 Descriptive statistics, mean (SD)

	MSLT group ($n=38$)	OSA Group ($n=172-178$)
Age (years)	43.2 (10.8)	55.5 (10.5)
Female sex, number (%)	12 (32)	43 (24)
Male sex, number (%)	26 (68)	135 (76)
Height, cm	176.4 (8.6)	176.5 (9.1)
Weight, kg	84.5 (17.6)	97.3 (19.2)
BMI, kg m^2	27.0 (4.8)	31.1 (5.4)
Education, number (%)		
Up to 10 years (elementary and middle school)	8 (21)	65 (36)
Started (1–2 years) or finished high school	19 (50)	51 (29)
College or university degree < 4 years	7 (18)	41 (23)
College or university degree > 4 years	4 (11)	18 (10)
Work situation		
Homemaker	1 (3)	4 (2)
Disability pension	8 (21)	42 (24)
Retired	0 (0)	30 (17)
Sick leave	8 (21)	7 (4)
Part time job	4 (11)	11 (6)
Full time job/student	17 (45)	82 (46)
Smoking status, number (%)		
Never smoked	16 (42) ^a	69 (39) ^c
Previous smoker	9 (24) ^a	70 (39) ^c
Current smoker	12 (32)	40 (23)
Alcohol consumption		
Never	1 (3) ^b	24 (14)
Rarely	24 (63) ^b	93 (52)
Weekly	10 (26) ^b	56 (32)
Daily	1 (3) ^b	3 (2)

^a $n=25$

^b $n=36$

^c $n=13$

Table 2 Test–retest reliability of Epworth sleepiness scale items ($n=50$)

ESS item	Kw	95% CI
Sitting and reading	0.78	0.50–1.06
Watching TV	0.76	0.49–1.04
Sitting inactive in a public place (e.g., a theater or a meeting)	0.79	0.52–1.07
As a passenger in a car for an hour without a break	0.70	0.43–0.97
Lying down to rest in the afternoon when circumstances permit	0.80	0.53–1.07
Sitting and talking to someone	0.61	0.34–0.87
Sitting quietly after a lunch without alcohol	0.80	0.53–1.08
In a car, while stopped for a few minutes in traffic	0.75	0.47–1.02
ESS total score	0.81 ^a	0.65–0.97

ESS Epworth sleepiness scale, Kw weighted kappa, CI confidence interval

^aIntraclass correlation coefficient

In the MSLT group ($n=36$), there were statistical differences in six of the eight items scores and the ESS total scores between those reporting sleeping during the day compared to non-sleepers. In contrast, there was a difference only on one ESS item when dividing groups according to the presence of snoring or sleepiness while driving (Table 4).

In the larger sample being evaluated for OSA ($n=178$), there were differences in item scores and the ESS total score among those reporting snoring or not. However, there were no differences in scores between groups divided according to the median AHI index or according to being in possession of a CPAP machine (Table 5).

Discussion

In this study, we investigated the reliability and validity of a Norwegian version of the ESS. The internal consistency reliability was acceptable, in line with recommendations with Cronbach's $\alpha > 0.80$ [21]. Assessment of the test–retest reliability among patients evaluated for OSA showed good agreement and therefore supports the reliability of the Norwegian ESS. We found only a fair to moderate association of the different ESS items and total score with the number of times a patient fell asleep and mean sleep latency found on the MSLT, mainly in a subset of patients who had a total ESS score > 10 . These findings may question the validity of the ESS when using MSLT variables as a gold standard for daytime sleepiness.

Our finding of a good test–retest reliability of the Norwegian ESS is comparable to that of other translations of the ESS [3, 12, 22, 23]. In contrast, a recently published study, where the ESS was administered to 142 patients with sleep apnea syndrome twice with an average time interval of 71 days, reported highly variable ESS scores over time [24]. This larger variability may be related to the longer time span between the assessments than in the present study.

The ESS represents a subjectively reported situational sleep propensity [25], and there are conflicting results about the association of the ESS with other objective measures of sleep propensity. Some previous studies have shown that higher ESS scores are associated with lower mean sleep latencies of the MSLT [2, 3]. In the present study, we found only fair to moderate associations between the MSLT and different ESS items and total ESS score in patients complaining of EDS, in line with previous research [5, 8–10] that have questioned whether the ESS can replace objective measures of sleep propensity like the MSLT [8–

Table 3 Spearman's rank correlation of the Epworth sleepiness scale (ESS) item and total score with the multiple sleep latency test (MSLT)

	Total sample ($n=37$)		Subset with ESS > 10 ($n=25$)	
	Number of times patient fell asleep	Mean latency to sleep	Number of times patient fell asleep	Mean latency to sleep
Sitting and reading	0.00	0.16	-0.24	-0.00
Watching TV	0.13	0.10	-0.41*	0.14
Sitting inactive in a public place (e.g. a theatre or a meeting)	0.18	0.12	-0.05	0.09
As a passenger in a car for an hour without a break	0.18	-0.04	0.04	-0.14
Lying down to rest in the afternoon when circumstances permit	0.43**	0.09	0.29	-0.31
Sitting and talking to someone	0.03	-0.17	-0.17	-0.37
Sitting quietly after a lunch without alcohol	0.35*	-0.10	0.26	-0.31
In a car, while stopped for a few minutes in traffic	-0.06	-0.05	-0.29	-0.18
ESS total score	0.09	-0.09	-0.23	-0.24

* $p < 0.05$ (two-tailed), ** $p < 0.01$ (two-tailed)

Table 4 ESS items and total scores according to presence or absence of snoring, sleeping during the day, and sleepiness while driving in the MSLT group, mean (SD)

	Snoring			Sleeping during the day			Sleepiness while driving		
	Yes	No	<i>p</i> value	Yes	No	<i>p</i> value	Yes	No	<i>p</i> value
<i>n</i>	19	17		23	14		8	29	
Sitting and reading	2.3 (0.8)	1.7 (1.0)	0.08	2.1 (0.9)	1.7 (1.0)	0.20	2.1 (0.8)	1.9 (1.0)	0.62
Watching TV	2.5 (0.8)	1.8 (1.1)	0.02	2.5 (0.8)	1.6 (1.1)	0.01	2.6 (0.5)	2.0 (1.0)	0.14
Sitting inactive in a public place (e.g. a theatre or a meeting)	2.0 (0.9)	1.3 (1.3)	0.08	2.0 (1.0)	0.9 (1.1)	0.00	1.9 (1.0)	1.5 (1.2)	0.44
As a passenger in a car for an hour without a break	1.8 (0.8)	1.8 (1.2)	0.92	2.3 (0.9)	1.1 (0.8)	0.00	2.6 (0.5)	1.6 (1.0)	0.01
Lying down to rest in the afternoon when circumstances permit	2.8 (0.7)	2.6 (0.7)	0.55	2.9 (0.6)	2.5 (0.8)	0.12	3.0 (0.0)	2.7 (0.8)	0.22
Sitting and talking to someone	0.6 (0.7)	0.6 (0.9)	0.87	1.0 (0.8)	0.0 (.00)	0.00	1.1 (0.8)	0.4 (0.7)	0.02
Sitting quietly after a lunch without alcohol	1.4 (0.8)	0.9 (1.0)	0.16	1.6 (0.7)	0.5 (0.8)	0.00	1.5 (0.9)	1.1 (0.9)	0.23
In a car, while stopped for a few minutes in traffic	0.6 (0.7)	0.3 (0.6)	0.20	0.7 (0.7)	0.1 (0.3)	0.01	0.6 (0.7)	0.4 (0.6)	0.35
ESS total score	13.9 (4.2)	11.6 (5.7)	0.18	14.9 (4.2)	8.9 (3.9)	0.00	15.5 (3.8)	11.9 (5.1)	0.08

10]. Some of the prior studies go on to suggest that the ESS and MSLT may evaluate different aspects of sleepiness and that the use of both methods can be complementary to each other. This seems reasonable, as MSLT is a time-consuming, burdensome, and relatively expensive test to perform

when compared to the relative ease and inexpensive administration of the ESS. The MSLT is a commonly used method to objectively measure sleep propensity and is a valuable aid in the diagnosis of illnesses such as narcolepsy. The ESS cannot replace MSLT in these situations. The

Table 5 ESS item and total scores according to snoring status, $AHI \leq 24 \text{ h}^{-1}$ vs. $AHI > 24 \text{ h}^{-1}$, possession of a CPAP machine in the OSA group, Mean (S.D.)

	Snoring			Apnea–Hypopnea Index			Possession of a CPAP machine		
	Yes	No	<i>p</i> value	$AHI > 24 \text{ h}^{-1}$	$AHI \leq 24 \text{ h}^{-1}$	<i>p</i> value	Yes	No	<i>p</i> value
<i>n</i>	81–83	47–48		76–79	67–68		126–128	27–29	
Sitting and reading	1.6 (1.0)	1.1 (1.1)	0.02	1.5 (1.0)	1.4 (1.0)	0.68	1.5 (1.1)	1.3 (0.9)	0.37
Watching TV	1.9 (0.9)	1.5 (1.0)	0.03	1.8 (0.9)	1.8 (0.9)	0.75	1.8 (0.9)	1.5 (0.7)	0.09
Sitting inactive in a public place (e.g. a theatre or a meeting)	1.1 (1.0)	1.0 (0.9)	0.55	1.2 (1.0)	1.1 (0.9)	0.66	1.2 (1.0)	1.0 (1.0)	0.45
As a passenger in a car for an hour without a break	1.3 (1.2)	1.1 (1.0)	0.35	1.3 (1.1)	1.2 (1.1)	0.46	1.3 (1.1)	1.1 (1.0)	0.48
Lying down to rest in the afternoon when circumstances permit	2.6 (0.7)	2.5 (0.8)	0.30	2.7 (0.6)	2.5 (0.7)	0.13	2.6 (0.7)	2.5 (0.9)	0.53
Sitting and talking to someone	0.4 (0.7)	0.3 (0.6)	0.22	0.4 (0.6)	0.4 (0.7)	0.73	0.4 (0.7)	0.3 (0.5)	0.68
Sitting quietly after a lunch without alcohol	1.0 (1.0)	0.7 (0.9)	0.08	0.9 (0.9)	1.0 (1.0)	0.33	0.9 (1.0)	0.8 (0.8)	0.54
In a car, while stopped for a few minutes in traffic	0.4 (0.8)	0.1 (0.4)	0.02	0.4 (0.7)	0.3 (0.6)	0.43	0.4 (0.7)	0.2 (0.4)	0.29
ESS total score	10.4 (5.0)	8.4 (4.8)	0.03	10.2 (4.7)	9.7 (5.0)	0.57	10.0 (5.1)	8.9 (4.0)	0.25

findings in the present study suggest that the use of the ESS alone to quantify and reliably diagnose EDS is not sufficient, and we believe the MSLT is still needed to objectively quantify sleep propensity.

Some limitations of the study should be noted. The MSLT was performed during the day before an overnight PSG. This practice varies from the accepted convention of performing the PSG the night before the MSLT [16, 17], but because our laboratory performs these studies in an ambulant setting, recording the PSG the night of the MSLT makes it more cost-effective, efficient, and convenient for both the patient and the laboratory. This method of performing MSLT in an ambulatory setting is largely based on almost 30 years of tradition at our laboratory. Guidelines defining the use of the MSLT have changed over the years, where the newest guidelines published in 2005 state that “the MSLT must be performed immediately following PSG recorded during the individual’s major sleep period” [17]. Previous guidelines however have not been that categorical, for example, stating that “the MSLT is generally performed on a day following a clinical PSG recording” [26]. We have continued our practice of performing PSGs after MSLTs mainly due to logistic reasons. The obvious and biggest disadvantage of our current practice is the lack of objective documentation of sleep quality the night before the MSLT. We are unaware of any studies that have previously validated any form for ambulatory MSLT against the “gold standard” in-patient MSLT, and we are therefore in the process of initiating such a validation study.

The Norwegian version of the ESS used in this study was translated at Akershus University Hospital in 2001, when we were unaware that other researches worked on another Norwegian translation of the same scale. This version of the scale has been used in a telephone survey [27]; however, we are not aware of assessment of its psychometric properties. The two versions differ only slightly from one another, most significantly in the translation of the English verb dozing, where the two versions have selected two distinct Norwegian verbs. Another difference is that in the version used in this study, the ESS items are, as in the original version, translated in the present participle (i.e., sitting and reading), while the other version uses the infinitive translation of the verbs (i.e., to sit and read). Interestingly, in one study, where the aim was to develop a Greek version of the ESS, the researchers translated an English ESS, where instead of the verb dozing, probability of sleepiness and sleeping was used [23].

The present study was conducted in two different populations, one consisting of patients evaluated for OSA and the other of patients with a sleep disturbance, who were referred to a sleep study and an MSLT. This study documents the psychometric properties of the Norwegian version of the ESS, which we therefore think can be used in practice. However, the Norwegian version has the same limitations as other

language versions of the ESS. The measurement properties of the ESS clearly vary according to the population in which it is used, which has been emphasized in previous reports [7, 28]. This noted that the ESS is still an important questionnaire for the assessment of subjective sleep propensity, especially in patients who are evaluated for or have been diagnosed with OSA. It is translated into several different languages and used in many countries around the world [2, 11, 12, 22, 23]. The ESS has the advantage of evaluating average sleep propensity in recent times [29], while the MSLT evaluates the condition of the patient in a very specific situation and does not necessarily reflect daily life [22]. We conclude that the Norwegian version of the ESS had acceptable internal consistency and test–retest reliability. The association of the ESS items and total score with the MSLT was only fair to moderate, in line with previous reports from other countries.

References

1. Johns MW (1991) A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 14:540–545
2. Vignatelli L, Plazzi G, Barbato A et al (2003) Italian version of the Epworth sleepiness scale: external validity. *Neurological Sciences* 23:295–300
3. Chung KF (2000) Use of the Epworth Sleepiness Scale in Chinese patients with obstructive sleep apnea and normal hospital employees. *J Psychosom Res* 49:367–372
4. Chervin RD, Aldrich MS, Pickett R, Guilleminault C (1997) Comparison of the results of the Epworth Sleepiness Scale and the Multiple Sleep Latency Test. *J Psychosom Res* 42:145–155
5. Olson LG, Cole MF, Ambrogetti A (1998) Correlations among Epworth Sleepiness Scale scores, multiple sleep latency tests and psychological symptoms. *J Sleep Res* 7:248–253
6. Fong SY, Ho CK, Wing YK (2005) Comparing MSLT and ESS in the measurement of excessive daytime sleepiness in obstructive sleep apnoea syndrome. *J Psychosom Res* 58:55–60
7. Johns MW (1994) Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep* 17:703–710
8. Chervin RD, Aldrich MS (1999) The Epworth Sleepiness Scale may not reflect objective measures of sleepiness or sleep apnea. *Neurology* 52:125–131
9. 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. *Psychiatr Clin Neurosci* 53:301–302
10. Benbadis SR, Mascha E, Perry MC, Wolgamuth BR, Smolley LA, Dinner DS (1999) Association between the Epworth sleepiness scale and the multiple sleep latency test in a clinical population. *Ann Intern Med* 130:289–292
11. Chen NH, Johns MW, Li HY et al (2002) Validation of a Chinese version of the Epworth sleepiness scale. *Qual Life Res* 11:817–821
12. Izci B, Ardic S, Firat H, Sahin A, Altinors M, Karacan I (2007) Reliability and validity studies of the Turkish version of the Epworth Sleepiness Scale. *Sleep Breath* 12:161–168
13. Stavem K, Kjelsberg FN, Ruud EA (2004) Reliability and validity of the Norwegian version of the Functional Outcomes of Sleep Questionnaire. *Quality of Life Research* 13:541–549
14. 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* 9:5–11
15. AASM Task Force (1999) 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* 22:667–689
 16. Thorpy MJ (1992) The clinical use of the Multiple Sleep Latency Test. The Standards of Practice Committee of the American Sleep Disorders Association. *Sleep* 15:268–276
 17. Littner MR, Kushida C, Wise M et al (2005) Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep* 28:113–121
 18. Rechtschaffen A, Kales A (1968) A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. NIH Publication No. 204. Washington, DC
 19. Kundel HL, Polansky M (2003) Measurement of observer agreement. *Radiology* 228:303–308
 20. Altman DG (1991) *Practical Statistics for Medical Research*. Chapman and Hall, London
 21. Nunnally JC, Bernstein IH (1994) *Psychometric theory*. McGraw-Hill, New York
 22. Bloch KE, Schoch OD, Zhang JN, Russi EW (1999) German version of the Epworth Sleepiness Scale. *Respiration* 66:440–447
 23. Tsara V, Serasli E, Amfilochiou A, Constantinidis T, Christaki P (2004) Greek version of the Epworth Sleepiness Scale. *Sleep Breath* 8:91–95
 24. Nguyen AT, Baltzan MA, Small D, Wolkove N, Guillon S, Palayew M (2006) Clinical reproducibility of the Epworth Sleepiness Scale. *J Clin Sleep Med* 2:170–174
 25. Johns MW (2002) Sleep propensity varies with behaviour and the situation in which it is measured: the concept of somnificity. *J Sleep Res* 11:61–67
 26. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S (1986) Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep* 9:519–524
 27. Pallesen S, Nordhus IH, Omvik S, Sivertsen B, Tell GS, Bjorvatn B (2007) Prevalence and risk factors of subjective sleepiness in the general adult population. *Sleep* 30:619–624
 28. Kim H, Young T (2005) Subjective daytime sleepiness: dimensions and correlates in the general population. *Sleep* 28:625–634
 29. Johns MW (1992) Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep* 15:376–381