Sleep, 15(4):376–381

Reliability and Factor Analysis of the Epworth Sleepiness Scale

Murray W. Johns

Sleep Disorders Unit, Epworth Hospital, Melbourne, Victoria, Australia

Summary: The Epworth Sleepiness Scale (ESS) is a self-administered eight-item questionnaire that has been proposed as a simple method for measuring daytime sleepiness in adults. This investigation was concerned with the reliability and internal consistency of the ESS. When 87 healthy medical students were tested and retested 5 months later, their paired ESS scores did not change significantly and were highly correlated ($r = 0.82$). By contrast, ESS scores that were initially high in 54 patients suffering from obstructive sleep apnea syndrome returned to more normal levels, as expected, after 3–9 months' treatment with nasal continuous positive airway pressure. The questionnaire had a high level of internal consistency as measured by Cronbach's alpha (0.88). Factor analysis of item scores showed that the ESS had only one factor for 104 medical students and for 150 patients with various sleep disorders. The ESS is a simple and reliable method for measuring persistent daytime sleepiness in adults. **Key Words:** Sleepiness—Epworth—Questionnaire—Factor analysis—Reliability.

In the assessment of sleep habits and sleep disorders, a subject's general level of sleepiness during the day is an important characteristic that should be measured routinely. The Multiple Sleep Latency Test (MSLT) is widely used to measure daytime sleepiness, in the sense of the propensity to fall asleep when encouraged to do so in a nonstimulating environment (1). However, the time-consuming nature and expense of the MSLT are such that, all too often, such tests have not been done. There is great need for a simpler alternative.

Some earlier attempts to quantify daytime sleepiness on the basis of the subject's responses to one or more questions succeeded to some extent (2,3). However, each attempt involved different questions for which normative data have seldom been available. The lack of standardization has prevented comparisons being made between different studies. By contrast, other investigations have failed to distinguish different levels of daytime sleepiness on the basis of subjective reports (4–6). One can but speculate about the reasons for this failure, but the methods were not validated and the questions asked or the grading of responses cannot have been appropriate. The Stanford Sleepiness Scale (SSS) measures feelings of sleepiness or, perhaps more accurately, of tiredness at a particular time (7). However, the SSS has been found to be quite unsatisfactory when assessing sleep propensity as measured by the MSLT (8,9). A Sleep–Wake Activity Inventory (SWAI) has recently been developed with 35 items (10). Subjects respond to each item by marking a visual analogue scale at an appropriate position along its 10-cm length. In a factor analysis of item scores there were two factors, one of which was related to daytime sleepiness as measured by the MSLT, the other to "psychological distress". This supports the contention that appropriate questions in a standardized questionnaire can provide clinically useful measurements of a subject's general level of daytime sleepiness.

The Epworth Sleepiness Scale (ESS) has been proposed specifically for that purpose, as a simple method for measuring the general level of daytime sleepiness or sleep propensity in adults (11). The ESS is a brief, self-administered questionnaire that asks the subject to rate on a scale of 0–3 the chances that, over "recent times", he would have dozed in eight specific situations that are commonly met in daily life (0 = would never doze; 3 = high chance of dozing). Thus, the subject is asked to characterize, retrospectively, part of his usual behavior in a variety of situations that are more or less soporific. Subjects are asked to distinguish dozing behavior from feelings of tiredness. The ESS score is the sum of eight item-scores and can range from 0 to 24.

Evidence about the validity of ESS measurements has been reported previously (11). In patients with various sleep disorders ESS scores were correlated sig-

Accepted for publication April 1992.
Address correspondence and reprint requests to Murray W. Johns, Sleep Disorders Unit, Epworth Hospital, Melbourne, Victoria 3121, Australia.

nificantly with mean sleep latencies measured in MSLTs ($r = -0.514$, n = 27, $p < 0.01$).

Scores on the ESS significantly distinguished control subjects from several groups of patients with sleep disorders that are known to be associated with different levels of daytime sleepiness. In patients with obstructive sleep apnea syndrome (OSAS) ESS scores were related to the severity of that disorder as measured by the respiratory disturbance index and by the minimum arterial oxygen saturation measured overnight during apneas. Patients with chronic psychophysiological or idiopathic insomnia tend to have very low ESS scores, indicative of a low sleep propensity.

The present investigation was concerned, first, with the test–retest reliability of ESS scores and, second, with the internal structure and consistency of the questionnaire. Healthy medical students are shown here to have a wide range of ESS scores that did not change significantly over a 5-month period. By contrast, the high ESS scores indicative of excessive daytime sleepiness in patients with OSAS are shown to return to more normal levels, as expected, after a few months' treatment with nasal continuous positive airway pressure (CPAP). The internal consistency of the ESS questionnaire is demonstrated by item analysis in two different groups of subjects—medical students and patients with a variety of sleep disorders. Factor analysis of ESS item scores, performed separately in these same two groups of subjects, shows that the questionnaire measures only one main variable.

These results, in addition to those reported earlier (11), provide evidence that the ESS is a simple and reliable method for measuring the general level of daytime sleepiness in adults.

METHODS

Details of the ESS questionnaire have been published previously (11).

Subjects

All 104 third-year medical students who had assembled for a teaching session at Monash University Medical School, Melbourne, answered the ESS at the same time, without discussion. Their mean age was 20.9 ± 2.8 (SD) years. There were 55 men and 49 women, all of whom were ostensibly healthy although no attempt was made to investigate their general health or sleep habits in detail. The sleep habits of similar medical students at Monash University have been described previously using a sleep questionnaire (12). The mean duration of sleep on week nights was 7.7 ± 0.7 (SD) hours, increasing to 8.4 ± 1.2 (SD) hours on weekends, figures that do not suggest chronic sleep deprivation, at least for the majority of such students. Apart from sleeping about half an hour less per day by waking earlier in the morning in summer than in winter, most aspects of the sleep habits of such students do not change significantly over a year (12).

During the academic year in question these students were not involved with night calls. Their daytime sleepiness was not measured by means other than the ESS. It was assumed that although the daytime sleepiness of a few may increase or decrease over a period of months because of changes in their sleep habits, the sleepiness of most, and of the group as a whole, would remain constant, as others have found using MSLTs (13).

The students in the present investigation first answered the ESS at the beginning of May 1991, 2 months after the start of their academic year. The same class of students was asked, without warning, to answer the ESS a second time at the end of September, 5 months after their first response and 2 months after their winter vacation. Eighty-seven students were identified as having completed the questionnaire on both occasions. The ESS scores of these 87 did not differ significantly from those of the whole group of 104. The 87 paired ESS scores were used to assess the test–retest reliability of the questionnaire.

Over a 9-month period in the Sleep Disorders Unit at Epworth Hospital there were 54 patients with OSAS who were treated successfully with nasal CPAP for at least 3 months and who had answered the ESS before and after treatment. Their mean age was 53.1 ± 11.0 (SD) years with a range from 28 to 78 years. Their respiratory disturbance index (RDI) was defined as the number of apneas and hypopneas causing a fall in arterial oxygen saturation of at least 3% per hour of sleep. Apneas involved cessation of nasal and oral airflow, and hypopneas at least a 50% reduction in airflow for 10 seconds or more. The polysomnographic methods have been described previously (14). The mean RDI before treatment was 27.1 ± 15.5 (SD) (range 4–56). The minimum arterial oxygen saturation recorded for each patient during apneas overnight (before treatment) had a mean of 73 ± 10.5% (SD) (range 50–92%).

The optimum nasal CPAP pressure required to control snoring and apneas and to maintain the SaO_2 at ≥90% was determined for each patient during a second night's polysomnography. That pressure was then fixed in a CPAP pump that was used each night at home. Only patients who reported compliance with this treatment were included here.

The ESS scores of 150 consecutive patients with various sleep disorders have been reported previously (11). Their diagnoses included OSAS, primary snoring, narcolepsy, idiopathic hypersomnia, periodic limb

movement disorder and idiopathic or psychophysiological insomnia. Ages ranged from 18 to 78 years. Half of the patients with OSAS who are treated by nasal CPAP, as above, had their pretreatment ESS scores included in this group, which was finalized before the CPAP treatment group. These 150 patients' scores on each item of the ESS were submitted to factor analysis. A similar analysis was performed separately on the item scores of the 104 medical students.

Statistical methods

All statistical analysis was done on a personal computer using the commercially available package of programs called Statistica® (Statsoft Inc., Tulsa, OK). The tests included Student's t tests for dependent or independent samples, Pearson correlation coefficients and the measurement of skewness and kurtosis of an approximately normal distribution of ESS scores. Differences between the frequency distributions of item scores were tested by χ^2 tests (df = 3: no Yates correction). Statistical significance was accepted at $p < 0.05$ in two-tailed tests. The program "Reliability" was used for item analysis and the calculation of Cronbach's statistic, alpha, which gives a measure of the internal consistency of the questionnaire (15). Factor analysis of ESS item-scores was performed by the program "Factor". Communalities were calculated by multiple regression rather than being estimated from variances. As is usual, only factors with eigenvalues >1.0 were retained. Varimax or other rotation was not appropriate because there was only one main factor in each of the two separate factor analyses.

It is feasible to perform factor analysis on ordinal (ranked) data such as ESS item-scores that can only be 0, 1, 2 or 3 (16). However, such analysis assumes that the variables are normally distributed. In fact, the distribution of item-scores was significantly different from normal for all items except item 5 for the patients and all except items 5, 6 and 8 for the students ($p < 0.001$, tested by separate χ^2 tests, df = 3). It was not feasible to normalize the data separately for each item, and it was accepted that there may be some limitation in the accuracy of calculations as a result. If this were a practical problem it would presumably become apparent in the comparison of results from the two populations in which item-scores were distributed quite differently (see below).

RESULTS

ESS scores of medical students

The mean ESS score for the whole group of 104 medical students was 7.6 ± 3.9 (SD). The range was 0–18 and the modal score was 7 (Fig. 1). The distribution was approximately normal, but with a skewness of 0.32 and kurtosis of −0.53. These scores were slightly higher than those of 30 normal control subjects in a previous investigation in whom the mean ESS score was 5.9 ± 2.2 (SD), the modal score was 6 and the range was 2–10 (2). The difference was statistically significant ($t = 2.37$, df = 132, $p < 0.02$).

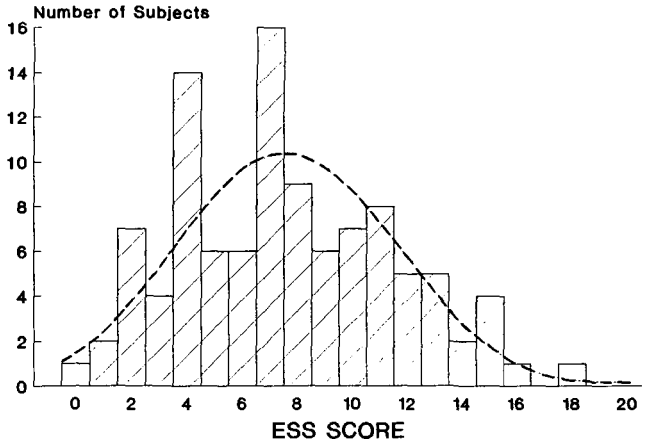

FIG. 1. The distribution of ESS scores from 104 medical students and the normal distribution (dotted).

Test-retest reliability

Comparisons between the ESS responses of 87 students on two separate occasions, 5 months apart, enabled an estimate to be made of the test–retest reliability of the questionnaire. On the first occasion their mean ESS score was 7.4 ± 3.9 (SD) and on the second occasion was 7.6 ± 3.8 (SD). The mean difference between these paired scores was 0.20 ± 2.3 (SD), which was not statistically significant ($t = 0.79$, df = 86, $p = 0.43$). The paired scores differed by no more than 1 in 51.7% of students, by no more than 2 in 81.6% and by no more than 4 in 96.6%. The Pearson correlation coefficient between the 87 paired scores was 0.822 ($p < 0.001$).

By contrast, the 54 patients with OSAS who were treated by nasal CPAP had ESS scores of 14.3 ± 3.6 (SD) with a range of 5–21 before treatment and 7.4 ± 4.1 (SD) with a range of 0–16 after treatment. The mean difference between these paired scores was 7.0 ± 5.2 (SD), which was statistically significant ($t = -9.59$, df = 53, $p < 0.001$).

Thus, ESS scores did not change significantly over several months under circumstances when daytime sleepiness was expected to remain constant but did change significantly as expected, after CPAP treatment for OSAS. These results indicate a reasonably high level of reliability for ESS scores.

TABLE 1. *The ESS item scores of (a) 40 sleepy patients and (b) 104 medical students and the significance of differences between those two groups for each item (χ^2 tests, 3 df)*

Item no.	Situation	Group	No. of subjects with item scores of 0	1	2	3	(p)
1	Sitting and reading	(a)	0	5	10	25	<0.001
		(b)	31	38	29	6	
2	Watching TV	(a)	0	5	7	28	<0.001
		(b)	41	46	12	5	
3	Sitting, inactive in a public place	(a)	1	8	19	12	=0.001
		(b)	40	38	18	8	
4	As a passenger in a car for an hour without a break	(a)	2	7	11	20	<0.001
		(b)	30	37	16	21	
5	Lying down to rest in the afternoon when circumstances permit	(a)	0	0	4	36	<0.001
		(b)	3	19	24	58	
6	Sitting and talking to someone	(a)	8	18	10	4	<0.001
		(b)	99	4	0	1	
7	Sitting quietly after a lunch without alcohol	(a)	0	10	12	18	<0.001
		(b)	38	38	20	8	
8	In a car, while stopped for a few minutes in the traffic	(a)	9	16	6	9	<0.001
		(b)	93	9	1	1	

Item analysis

Cronbach's statistic, alpha, was used as a measure of the internal consistency of the items in the questionnaire. This analysis was performed on item-scores from two separate groups of subjects—150 patients with various sleep disorders and 104 medical students. Alpha was 0.88 for the patients and 0.73 for the students. In neither group did the value of alpha increase after deleting any one of the items from the questionnaire. These results indicate a reasonably high level of consistency between the eight items of the ESS.

The eight situations described in the ESS are not all of the same soporific nature. Indeed, they were chosen on a priori grounds to be different. This is illustrated by comparing the item-scores of subjects in two different groups (Table 1). The first group was of 40 patients, described here as sleepy because they were all those with narcolepsy, idiopathic hypersomnia or severe OSAS (RDI > 30), selected from the group of 150 patients described previously (11). The mean ESS score for these 40 patients was 17.2 ± 3.7 (SD); much higher than for the 104 medical students (7.6 ± 3.9). These two groups also differed significantly in their item-scores for each item (Table 1). In all eight situations the 40 sleepy patients were much more likely than the students to doze. In both groups the situation in item 5 (lying down to rest in the afternoon when circumstances permit) was the most soporific. Every one of the 40 sleepy patients reported at least a moderate chance of dozing off then, of which 90% reported a high chance of doing so. However, more than three-quarters of the students also reported a moderate or a high chance of dozing in that situation. By contrast, the situation in items 6 and 8 (sitting and talking to someone, or in a car, while stopped for a few minutes in the traffic) were the least soporific for both groups. Thirty-five percent of the sleepy patients, compared with only 1% of students, reported at least a moderate chance of dozing when sitting and talking to someone. The other situations in items 1–4 and 7 were intermediate in their soporific nature for both groups.

TABLE 2. *The normalized factor loadings for items of the ESS reported by 150 patients and 104 students*

ESS item no.	Normalized factor loadings: Patients	Students
1	0.73	0.55
2	0.59	0.49
3	0.77	0.62
4	0.68	0.54
5	0.53	0.49
6	0.73	0.25
7	0.76	0.64
8	0.73	0.37

Factor analysis

Factor analysis was performed on the ESS item-scores of the 150 patients and the 104 students separately. This showed that there was only one main factor in each group (Table 2). The eigenvalue was 3.95 for the patients and 2.07 for the students. The normalized factor loadings were relatively high on all items for the patients, and on all but items 6 and 8 for the students. There were so few students dozing in those two situations (Table 1) that there was very little variance associated with those item-scores.

DISCUSSION

The results indicate, first, that the ESS questionnaire is reasonably reliable in the test–retest sense and, second, that it has a high level of internal consistency,

assessed by item analysis and factor analysis within two different groups of subjects.

ESS scores did not change significantly and were highly correlated when tested and retested 5 months later in students whose daytime sleepiness was expected to remain constant. By contrast, the high ESS scores of patients with OSAS, indicative of the excessive daytime sleepiness, which is a clinical feature of that disorder, were reduced to more normal scores, as expected, after 3 or more months' treatment with nasal CPAP.

The internal consistency of the ESS, measured by Cronbach's alpha, was high for the patients with a variety of sleep disorders (0.88) and a little lower, although still satisfactory (0.73), for the students. This consistency was not increased by deleting any of the items from either group. Factor analysis of item-scores showed that the ESS measured only one main variable within each group. Thus, the results support the use of the ESS as a sum-scale for which the derived measurement, the total ESS score, is simply the sum of eight item-scores. Scoring the results is therefore very quick and simple.

The ESS was designed to measure daytime sleepiness over the whole range, from very high to low levels. The items were chosen, therefore, to represent situations of a widely differing soporific nature. The results show that the relative soporific nature of those situations is about the same for very sleepy patients as it is for normal subjects. However, sleepy subjects doze more frequently and in more of those situations than do normal subjects.

Despite the similarities in the results from the patients and the students there were also minor differences. Cronbach's alpha and the normalized factor loadings indicated a level of internal consistency in responses to the ESS that was lower for the students than for patients. The reasons for this are not clear but may involve the nonnormal distribution of some item-scores or the fact that the range of ESS scores was higher among the patients (0–24) than the students (0–18).

The findings that ESS scores of the medical students were higher than those of healthy control subjects reported in a previous investigation (11) highlights some difficulties in establishing normal values. The control group in the previous investigation had only 30 subjects, all of whom were older than the present students. However, it may be that normal adults in their early 20s are sleepier, on average, than normal middle-aged adults, as their ESS scores suggest. There is corroborative evidence for this in MSLTs reported by others (17).

Nasal CPAP treatment for several months reduced the ESS scores of patients with OSAS to the same levels as those of the medical students, but still slightly higher than those of middle-aged, nonsnoring controls from the previous investigation, with whom comparison may be more appropriate. This difference did not quite reach statistical significance ($t = 1.89$, $df = 82$, $p = 0.06$). However, the daytime sleepiness of patients with OSAS, even after successful treatment with nasal CPAP, may remain slightly above normal, as indicated also by the results of Di Phillipo et al. (18) who used MSLTs to measure sleepiness under similar circumstances.

There was a wide range of ESS scores among the students and the patients here, as there had been within each diagnostic group of patients and the controls in the earlier investigation (11). At each end of the clinically normal range, particularly with scores greater than 9, there is a considerable overlap with scores that are abnormal in the sense that they are associated with sleep disorders that are known to affect daytime sleepiness. This overlap between normal and abnormal occurs also with daytime sleepiness measured by MSLTs. Some healthy adults fall asleep in less than 5 minutes, which is considered to represent pathological sleepiness in others (19).

When the ESS was developed it was intended that it should measure only that component of daytime sleepiness which persists from week to week and longer in a given subject, independent of changes with the time of day and from day to day. This is what the MSLT and its variants such as the Maintenance of Wakefulness Test (20) are believed to measure when the mean SL is determined from four to six naps on one day, so overcoming much of the time-of-day effect, and attempts have been made to normalize the preceding night's sleep and to remove the effects of drugs that would otherwise influence the results. This longer-term or persistent component of daytime sleepiness may be influenced by several factors such as a psychophysiological trait involving inherently different levels of sleep propensity in difficult subjects or the presence of one or more chronic sleep disorders such as obstructive sleep apnea, narcolepsy or periodic limb movement disorder. It was believed that this general level of daytime sleepiness would, over a period of time, become incorporated into each subject's way of life, modifying his chances of either dozing or staying awake when in various nonstimulating situations. It was assumed also that most adults would be able to report whether or not they dozed in those situations. The initial results with the ESS suggest that those suppositions were correct (11).

The ESS questionnaire is now shown to be reliable, to be internally consistent and to have only one main dimension in its variance. It is conceptually unique in measuring the whole range of sleep propensities, from very high to very low. It is simple and self-administered, taking a few minutes to complete and a few

seconds to score. With mounting evidence about its validity, it promises to provide a much-needed, simple, standardized method for measuring daytime sleepiness in adults. Nevertheless, the ESS cannot perform all the functions of the MSLT such as measuring changes in sleepiness from day to day or indicating the early onset of REM sleep, which is so important in the diagnosis of narcolepsy.

REFERENCES

1. Richardson G, Carskadon M, Flagg W, Van Dan Hoed J, Dement W, Mitler M. Excessive daytime sleepiness in man: multiple sleep latency measurements in narcoleptics vs. control subjects. *Electroencephalogr Clin Neurophysiol* 1978;45:621–7.
2. Sink J, Bliwise DL, Dement WC. Self-reported daytime somnolence and impaired respiration in sleep. *Chest* 1986;90(2): 177–80.
3. Schmidt-Nowara WW, Wiggins CL, Walsh JK, Bauer C. Prevalence of sleepiness in an adult population. *Sleep Res* 1989;18: 302.
4. Regestein QR, Ferber R, Johnson TS, Murawsk B, Strome M. Subjective sleepiness and vigilance in obstructive sleep apnea. *Sleep Res* 1987;16:414.
5. Crocker ID, Olson LG, Saunders NA, Hensley MJ, McKeon JL, Allen KM, Gyulay SG. Estimation of the probability of disturbed breathing during sleep before a sleep study. *Am Rev Respir Dis* 1990;142:14–8.
6. Jennum P, Sjol, A, Suadicani P. Predictive symptoms for snoring and obstructive sleep apnea. *Sleep Res* 1991;20A:329.
7. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. *Psychophysiology* 1973;10:431–6.
8. Pressman MR, Fry JM. Relationship of autonomic nervous system activity to daytime sleepiness and prior sleep. *Sleep* 1989; 12:239–45.
9. Cook Y, Schmitt F, Berry D, Gilmore R, Phillips B, Lamb D. The effects of nocturnal sleep, sleep disordered breathing and periodic movements of sleep on the objective and subjective assessment of daytime somnolence in healthy aged adults. *Sleep Res* 1988;17:95.
10. Rosenthal LD, Rosen A, Wittig RM, Zorik FJ, Roehrs TA, Roth T. A sleep–wake activity inventory to measure daytime sleepiness. *Sleep Res* 1991;20:130.
11. Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991;14:540–5.
12. Johns MW, Gay TJA, Goodyear MDE, Masterton JP. Sleep habits of healthy young adults: use of a sleep questionnaire. *Br J Prev Soc Med* 1971;25:236–41.
13. Zwyghuizen-Doorenbos A, Roehrs T, Schaeffer M, Roth T. Test–retest reliability of the MSLT. *Sleep* 1988;11:562–5.
14. Johns MW. Polysomnography at a sleep disorders unit in Melbourne. *Med J Aust* 1991;155:303–8.
15. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
16. Munro BH, Visintainer MA, Page EB. *Statistical methods for health care research.* Philadelphia: J.B. Lippincott, 1986.
17. Levine B, Roehrs T, Lamphere J, Zorick F, Stepanski E, Roth T. Daytime sleepiness in young adults. *Sleep Res* 1987;16:207.
18. Di Phillipo MA, Fry JM, Pressman MR. Objective measurement of daytime sleepiness following treatment of obstructive sleep apnea with nasal CPAP. *Sleep Res* 1988;17:167.
19. Johnson LC, Spinweber CL, Gomez SA, Matteson LT. Daytime sleepiness, performance, mood, nocturnal sleep: the effect of benzodiazepine and caffeine on their relationship. *Sleep* 1990; 13(2):121–35.
20. Mitler M, Gujavarty KS, Browman CP. Maintenance of wakefulness test: a polysomnographic technique for evaluating treatment in patients with excessive somnolence. *Electroencephalogr Clin Neurophysiol* 1982;153:658–61.

Erratum

In the report from the ASDA entitled "The clinical use of the Multiple Sleep Latency Test" published in SLEEP volume 15(3), a few of the numbered references in the section "I. General" are incorrect. The correct numbers are as follows:

1. [1.0, 2.2]
2. [2.2]
3. [2.3, 2.4]
4. [2.3, 2.4, 2.5]
5. [2.4]