Sleep, 17(8):703–710

Sleepiness in Different Situations Measured by the Epworth Sleepiness Scale

Murray W. Johns

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

Summary: This investigation examines how the sleep propensity (SP) in one test situation, such as the Multiple Sleep Latency Test (MSLT), is related to sleepiness in daily life, as assessed by the Epworth Sleepiness Scale (ESS). This is a self-administered questionnaire, the item scores from which provide a new method for measuring SPs in eight different real-life situations. The ESS item scores were analyzed separately in four groups of subjects: 150 adult patients with a variety of sleep disorders, 87 medical students who answered the ESS on two occasions 5 months apart, 44 patients who also had MSLTs and 50 patients whose spouses also answered the ESS about their partner's sleepiness. The ESS item scores were shown to be reliable (mean rho = 0.56, $p < 0.001$). The SP measured by the MSLT was related to three of the eight item scores in a multiple regression ($r = 0.64$, $p < 0.001$). The results of nonparametric ANOVA, Spearman correlations, Wilcoxon's *t* tests, item and factor analysis suggest that individual measurements of SP involve three components of variation in addition to short-term changes over periods of hours or days: a general characteristic of the subject (his average SP), a general characteristic of the situation in which SP is measured (its soporific nature) and a third component that is specific for both subject and situation. The SP in one test situation, including the MSLT, may not be a reliable indicator of a subject's average SP in daily life. Perhaps we should reexamine the current concept of daytime sleepiness and its measurement. **Key Words:** Sleep propensity—Daytime sleepiness—Epworth Sleepiness Scale.

Excessive daytime sleepiness is an important symptom of several chronic sleep disorders, including obstructive sleep apnea (OSA), narcolepsy, idiopathic hypersomnia and periodic limb movement disorder. The Multiple Sleep Latency Test (MSLT) is widely accepted as the gold standard for measuring such sleepiness, a position recently reinforced by the American Sleep Disorders Association (1,2). The word sleepiness is used here in the sense of sleep propensity (SP), not of tiredness or other subjective feelings as measured by the Stanford Sleepiness Scale or a visual analog scale of alertness-sleepiness, which are not significantly related to SP (3,4).

It is generally assumed that if subjects are sleepy in the MSLT [i.e. they have a mean sleep latency (SL) <10 minutes, particularly if <5 minutes], they will also be sleepy in other "low-stimulus" situations. Conversely, if subjects are not sleepy in the MSLT (SL >10 minutes), they will not be sleepy in daily life either. It is assumed that the SP measured by the MSLT is a general characteristic of each subject. To what extent this is true is seldom questioned and has not been adequately tested.

Several variations of the MSLT have been proposed, the best known of which is the Maintenance of Wakefulness Test (MWT) (5). The MSLT is a more widely used and tested method than the MWT, but both have undergone tests of validity (2,5–8). Both are based on the reasonable premise that the quicker a subject falls asleep in the test situation the higher his level of sleepiness. Sangal et al. compared the results of MSLTs and MWTs done on the same day in 258 patients with a variety of sleep disorders (9). They found a correlation of 0.41 between the mean SLs measured in these two ostensibly valid tests of SP. This was statistically significant ($p < 0.001$), but accounted for less than 17% of the variance. Overall, about one-third of their patients in various diagnostic categories had discordant results, being relatively sleepy on one test but not on the other. The authors concluded that the MSLT was measuring sleepiness, but the MWT was measuring the "capacity for wakefulness". Because both tests measure the same variable, the latency before sleep onset, these results must raise doubts about the commonly accepted unitary concept of sleepiness.

Accepted for publication July 1994.
Address correspondence and reprint requests to Dr. Murray W. Johns, Sleep Disorders Unit, Epworth Hospital, Erin Street, Richmond, Victoria 3121, Australia.

The Epworth Sleepiness Scale (ESS) uses a quite different method for measuring SP based on the subject's retrospective reports of dozing behavior in eight situations that are commonly met in daily life (10). ESS scores have been correlated significantly with the mean SL in MSLTs (10). Like the MSLT and the MWT, ESS scores can distinguish groups of patients with various sleep disorders from normal subjects (10–12) and are correlated significantly with the severity of OSA measured by the frequency of apneas and hypopneas (11–15) in some but not all investigations. The ESS has a high test–retest reliability and a high level of internal consistency among its eight items (16).

This report examines the relationship between SPs measured in different situations. It is not intended to question the substantial body of evidence for the validity of the MSLT or related tests as measures of SP within their respective test situations. Rather, it is intended to examine the question of how far the sleepiness that such tests measure can be generalized to other situations and to daily life. To do this we must reexamine some current concepts and introduce some others.

The soporific nature of different situations and activities

It is self-evident that people fall asleep more quickly and more often if they lie down rather than stand up. Simply changing posture modifies SP, presumably by changing the input to the ascending reticular activating system from proprioceptors in the postural muscles and joints. When we purposely decide to sleep we initiate a process that decreases input to the ascending reticular activating system in several ways and thereby increases SP, no matter at what time of day or night this occurs. Reduced mental and physical activity in some situations during daily life can also increase SP involuntarily to the point where some subjects begin to doze, even when intending to stay awake. The extent to which any particular activity or situation facilitates dozing in most subjects can be called its relative soporific nature. This refers to the physical environment (e.g. room temperature, noise level) as well as the subject's posture, whether the eyes are open or closed, the subject's volition and motivation to stay awake rather than doze, and so on. For many but not all people, the MSLT test situation is about twice as soporific as the MWT test situation, that is, the mean SL in the MSLT is about half that in the MWT (9).

When groups of subjects are asked about their usual tendencies to doze in various situations in daily life, they often nominate watching television and reading as highly soporific, and attending a lecture or visiting friends as less so (17,18). The relative soporific nature of the eight situations in the ESS has been described previously for two groups of subjects, "sleepy" patients and medical students (16). The most soporific situations generally involve prolonged inactivity such as sitting or lying down with little body movement, little interaction with other people, particularly by talking, and little environmental stimulation by way of thermal discomfort, bright light, loud noise, etc. However, some subjects respond paradoxically and fall asleep in what seems like a high-stimulus situation, subjected to very loud noise and bright flashes of light (19). Many subjects respond to their first night in the sleep laboratory, with its novelty and attachment of wires, by having worse sleep than on subsequent nights in the laboratory or at home. Yet some, paradoxically, sleep better on their first night in the sleep laboratory than on later nights or at home (20). Similarly, some subjects fall asleep quicker (i.e. they have a higher SP) in the MWT when instructed to stay awake than in the MSLT when instructed to fall asleep (9). It should not surprise us, therefore, that despite overall similarities in the relative soporific nature of various situations we may also find differences between individual subjects or groups.

For the purposes of this discussion, the sometimes substantial but short-term fluctuations in SP, such as with the time of day, after taking a sedative drug or after a night's sleep deprivation, are ignored. The SP that a subject usually shows in a particular situation will be called his situational sleep propensity. The MSLT measures one situational SP, the MWT another. So, too, would each of the eight item scores of the ESS. These are less accurate than the measurements of SL in the MSLT and MWT, but cover a wider range of situations that are, arguably, more relevant to daily life. The ESS score is one way of combining several different situational SPs to provide a measure of a more general characteristic, the average SP in daily life.

This report examines the relationship between measurements of SP in the MSLT and each of the eight situations described in the ESS. There are four parts dealing with data collected from several groups of subjects. In the first part the SLs measured in MSLTs are compared with the ESS scores of the same patients. In the second, the reliability of measurements of situational SPs is examined. The third examines the relative soporific nature of situations in the ESS. The fourth describes the relationships between different situational SPs measured by ESS item scores in the same subjects.

METHODS

The ESS questionnaire and the method of its self-administration have been described previously (10). The different situations in the questionnaire, referred to by their item numbers, are described in Table 1. The subjects were in four groups, each used in one or more of the four parts of this investigation. The groups were as follows:

Group (i): Forty-four consecutive, fee-paying patients referred for investigation to the Sleep Disorders Unit of Epworth Hospital, Melbourne, who on clinical grounds alone were deemed necessary to have an MSLT after overnight polysomnography. They had four naps, at 10:00 a.m., 12:00 midday, 2:00 p.m. and 4:00 p.m. After investigation, 10 were found to have narcolepsy, 12 idiopathic hypersomnia, four periodic limb movement disorder ± restless legs syndrome, eight OSA, nine psychophysiological hypersomnia and one had hypersomnia after head injury. There were 24 men and 20 women with a mean age of 44.5 ± 14.0 (SD) years.

Group (ii): One hundred fifty consecutive patients at Epworth Hospital with a variety of sleep disorders who had been the subject of an earlier report (10). Thirty-two had primary snoring, 55 obstructive sleep apnea, 13 narcolepsy, 14 idiopathic hypersomnia, 18 psychophysiological insomnia and 18 had periodic limb movement disorder. There were 120 men and 30 women with a mean age of 45.8 ± 12.1 years.

Group (iii): Eighty-seven ostensibly healthy 3rd year medical students at Monash University, Melbourne, who had been part of an earlier investigation (16). They were all the students gathered at a teaching session who answered the ESS once and who, 5 months later and without warning, answered the ESS a second time. There were 46 male and 41 female students with a mean age of 20.6 ± 1.6 years.

Group (iv): Fifty consecutive patients among those referred to the Sleep Disorders Unit who happened to attend with their spouse. Most of these patients presented with suspected sleep apnea, their spouses complaining of their snoring, their "stopping breathing" at night and excessive sleepiness during the day. There were 44 male and six female patients with partners of the opposite sex. These patients' ages ranged from 28 to 79 years (mean = 55.1 ± 12.5 years). Patient and spouse were each asked to answer the ESS questionnaire independently, without assistance or discussion and while under supervision.

STATISTICAL METHODS

The ESS item scores were on a four-point ordinal scale (0–3). Kolmogorov-Smirnov tests showed that the distributions for many groups of data differed significantly from normal. Consequently nonparametric methods were used for the analyses (Spearman's correlation coefficients, Kendall's coefficient of concordance, Wilcoxon's t tests, Mann-Whitney U tests, χ^2 tests and Friedman's nonparametric ANOVA). Multiple regression analysis was used to determine the relationship between the mean SL in MSLTs (the distribution of which did not differ significantly from normal) and ESS item scores. Item analysis using Cronbach's statistic, alpha and factor analysis was performed on ESS item scores, separately for each group. If, as proved to be the case, the results were similar for each group, it would allay fears that the analyses might have been heavily influenced by differences in the distributions of item scores between items and also between groups of subjects. Statistical significance was accepted at $p < 0.05$ in two-tailed tests.

RESULTS

Sleep latency in the MSLT and ESS scores

For the patients in Group (i) there was a statistically significant correlation between the mean SL in the MSLT and the total ESS score (rho = −0.42, n = 44, $p < 0.01$) (Fig. 1). Of the 10 patients with narcolepsy-cataplexy eight had a mean SL in the MSLT ≤5 minutes, but one had a mean SL of 9 minutes and another of 12 minutes, as others have found occasionally with narcolepsy (2). However, all 10 narcoleptics here had ESS scores >12. Of all 16 patients who had a mean SL ≤5 minutes in the MSLT (regardless of diagnosis), 15 had an ESS score >12 and one had an ESS score of 10, the upper limit of normal. Likewise, 10 of the 11 patients with a mean SL of 5–10 minutes in the MSLT also had ESS scores >10, indicative of increased daytime sleepiness. Thus, most patients (92.6%) with a high SP in the MSLT also had increased SP according to their ESS scores. However, the reverse was not always so. Of the 40 patients with ESS scores >10, 25 (62.5%) were also excessively sleepy according to the MSLT, but 15 (37.5%) were not. The latter patients all complained of excessive daytime sleepiness and confirmed this in their high ESS scores, but they did not fall asleep in <10 minutes when under scrutiny in the MSLT.

When the mean SL in the MSLT for each patient was correlated separately with each of their eight ESS item scores, only three correlation coefficients were statistically significant (Table 1). Those subjects with the shortest SL in the MSLT were the most likely to doze "when sitting, inactive in a public place" (item 3), when "sitting quietly after lunch without alcohol" (item 7) and when "in a car, stopped for a few minutes in the traffic" (item 8). The tendency to doze in other situations, such as while watching television, was not significantly correlated with SL in the MSLT.

The multiple regression with all eight item scores as predictors of the mean SL in the MSLT was statistically significant ($r = 0.639$, $F = 3.02$, df = 8.35, p = 0.01)

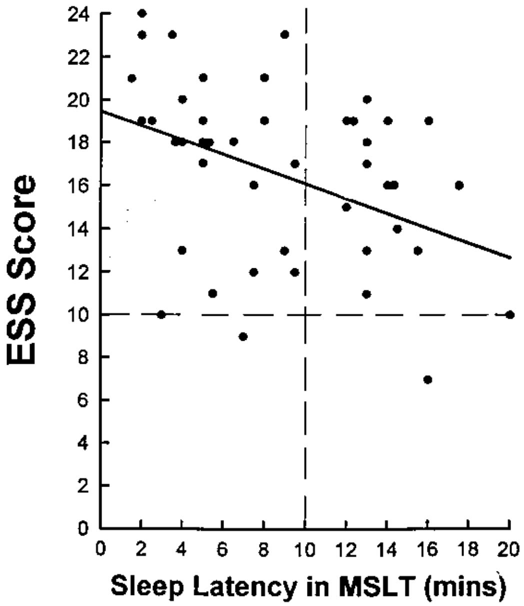

FIG. 1. Relationship between ESS scores and mean sleep latencies in MSLTs for 44 patients (rho = −0.42, p < 0.01).

TABLE 1. *Spearman correlations between ESS item scores and mean sleep latencies in MSLTs, and the multiple regression beta weights for ESS item scores as predictors of the mean sleep latency in the MSLT for 44 subjects*

ESS item no.	Situation	Mean sleep latency in MSLT			
		Rho	p	Beta weight	p
1	Sitting and reading	−0.23	ns	−0.20	ns
2	Watching TV	−0.19	ns	0.13	ns
3	Sitting, inactive in a public place (e.g. a theatre or a meeting)	−0.44	<0.01	−0.48	<0.01
4	As a passenger in a car for an hour without a break	0.00	ns	0.44	0.01
5	Lying down to rest in the afternoon when circumstances permit	−0.07	ns	0.01	ns
6	Sitting and talking to someone	−0.19	ns	0.14	ns
7	Sitting quietly after lunch without alcohol	−0.34	<0.05	−0.15	ns
8	In a car, while stopped for a few minutes in the traffic	−0.41	<0.01	−0.44	0.01

(Table 1). This regression explained 40% of the variance in MSLT results. Items 3 and 8 were significant independent predictors; item 7 was not. Instead, item 4 became a significant predictor through its correlations with other item scores; the direction of its prediction, however, was unexpected. The sleepiest patients by the MSLT were the least likely to doze when traveling "as a passenger in a car for an hour without a break". This needs confirmation before it could be considered to be other than a statistical aberration. The important point is that the SP of these patients in the MSLT was related significantly and, to some extent, independently to their tendency to doze in some ostensibly "low-stimulus" situations in daily life, but not in others. These results are at least consistent with the idea that SP depends partly on the situation in which it is measured.

Reliability of measurements of situational sleep propensity

Before discussing further the relationships between different situational SPs we should assess the reliability of such measurements, first in the MSLT and second in the ESS.

Reliability of SL measured in the MSLT

The overall mean SL for the patients in Group (i) was 8.8 ± 5.0 minutes. Friedman's ANOVA showed that the SLs did not differ significantly with the time of day (p > 0.2). All six within-subject correlations between the four SLs were statistically significant (mean rho = 0.62, p < 0.001). This represents a relatively high reliability for measurements of SL in the MSLTs, similar to that reported by others. For example, Sangal et al. (9) reported an equivalent mean r for SLs in both the MSLT and the MWT of 0.61 (n = 258, p < 0.001).

Reliability of ESS scores over 5 months

Responses by the medical students in Group (iii) to each item of the ESS were compared over a test–retest period of 5 months (Table 2). Friedman's ANOVA did not reveal any significant overall difference between these two sets of item scores. Nor did Wilcoxon's t test for dependent samples reveal any significant difference between the paired scores for any item. The matched pairs of item scores were all significantly correlated (mean rho = 0.56, p < 0.001). Within all 696 pairs of scores, 66.9% were identical and 95.1% differed by no more than one. The repeated total ESS scores were also correlated significantly (rho = 0.81, n = 87, p < 0.001).

Patient versus spouse ESS scores

The matched responses to the ESS items of the patients in Group (iv) and their spouses were compared (Table 2). Friedman's ANOVA did not reveal any significant overall difference between the patients' and spouses' item scores. However, Wilcoxon's t tests for dependent samples did show significant differences between their scores for items 4 and 5 (Table 2). The spouses may have slightly overestimated, or the patients may have underestimated, their situational SPs

TABLE 2. *Mean ESS item scores for four groups of subjects, with the significance of differences (Wilcoxon* t *tests) and Spearman correlations between repeated scores in medical students and between patient–spouse scores*

ESS item no.	Group (i) patients (n = 44)	Group (ii) patients (n = 150)	Group (iii)a students (n = 87)	Group (iii)b students (n = 87)	Difference between Groups (iii)a and (iii)b values	Correlation between Groups (iii)a and (iii)b rho values	p	Group (iv)c patients (n = 50)	Group (iv)d spouses (n = 50)	Difference between Groups (iv)c and (iv)d values	Correlation between Groups (iv)c and (iv)d rho values	p
1	2.45	1.58	1.08	1.16	>0.4	0.60	<0.001	1.94	2.08	>0.3	0.44	<0.001
2	2.48	1.78	0.77	0.86	>0.2	0.65	<0.001	2.44	2.48	>0.6	0.65	<0.001
3	2.07	1.18	0.85	1.00	>0.1	0.62	<0.001	1.44	1.56	>0.3	0.70	<0.001
4	2.34	1.06	1.30	1.29	>0.9	0.61	<0.001	1.79	2.08	<0.05	0.71	<0.001
5	2.95	2.27	2.26	2.29	>0.7	0.59	<0.001	2.56	2.76	<0.05	0.56	<0.001
6	1.18	0.51	0.03	0.03	>0.9	0.31	<0.001	0.66	0.72	>0.4	0.49	<0.001
7	2.05	1.23	0.97	0.93	>0.6	0.67	<0.001	1.74	1.90	>0.3	0.52	<0.001
8	1.14	0.56	0.14	0.07	>0.1	0.43	<0.001	0.52	0.50	>0.8	0.47	<0.001
Total of all ESS items	16.8	10.2	7.4	7.6	>0.4	0.81	<0.001	13.1	14.1	<0.05	0.74	<0.001

in those situations, whereas they did not consistently do so in other situations. A spouse may not often be present in the situation of item 5, "lying down to rest when circumstances permit", so their scores for that item may be less accurate than the patient's for want of direct observation. Nevertheless, the patient-spouse paired item scores were all significantly correlated (mean rho = 0.57, $p < 0.001$) (Table 2). Within all 400 pairs, 60.0% of scores were identical and 91.3% differed by no more than one. There was a small difference between the total ESS scores of these patients (mean = 13.1 ± 5.2) and their spouses (mean = 14.1 ± 5.2). This just reached statistical significance by Wilcoxon's *t* test (p = 0.04). Nevertheless, the paired total ESS scores were highly correlated (rho = 0.74, n = 50, $p < 0.001$) (Table 2).

These results indicate that, at least for most of the situations in the ESS, spouses can give reliable assessments of their partners' dozing behavior. Such behavior may well have been an earlier topic of conversation between them, but this is unlikely to have involved the specific details reported later in the ESS. The results suggest rather that ESS scores reflect independently observable behavior and not simply subjective feelings of sleepiness.

In summary, the reliability of ESS item scores for a particular subject is high whether repeated by the same subject at different times or by another person such as a spouse. This reliability is only slightly less than that for measurements of SL on the same day in the MSLT. As expected, the reliability of total ESS scores is higher again than for the separate item scores.

Relative soporific nature of situations in the ESS

Within any group of subjects, the mean score for each of the eight ESS item scores gives a measure for those subjects of the relative soporific nature of the different situations (Table 2). Within each of the four groups of subjects here Friedman's ANOVA demonstrated significant overall differences in the item scores between situations ($p < 0.001$ for each group).

The mean item scores were ranked within each group, enabling comparisons to be made between the relative soporific nature of the ESS situations while overcoming differences in sleepiness between the groups (Table 3). The rankings were very similar for three of the four groups. In the three groups of patients, items 5, 2 and 1 were the most soporific situations, items 6 and 8 the least soporific, and items 3, 4 and 7 were intermediate. Unlike the three groups of patients, the ranking for the students was different in that they reported that traveling "as a passenger in a car for an hour without a break" (item 4) was more soporific for them than "watching TV" (item 2). This serves to emphasize that what appears to be the same situation can have quite disparate effects on different subjects' SP. This is presumably because of subtle differences in what the situation means to each subject. Thus, there may be some degree of specificity due to subject × situation interactions in measurements of SP that cannot be accurately predicted from measurements in other subjects or other situations.

Relationships between SPs in different situations for the same subject

Within-subject comparisons of ESS item scores provide a new opportunity to assess sleepiness in eight different situations in the same subject. The item scores within each of the four groups of subjects were analyzed separately so that some idea could be gained of the constancy of the relationships. The analysis was in three stages: first, by Spearman's correlation coefficients between all 28 pairs of item scores within sub-

TABLE 3. *Situations in the ESS (identified by their item numbers) ranked according to their relative soporific nature in four groups of subjects*

			Situations (ESS item nos.)							
		n	Most sopo-rific							Least sopo-rific
Group (i)	patients	150	5	2	1	7	3	4	8	6
Group (ii)	students	87	5	4	1	7	3	2	8	6
Group (iii)	patients	50	5	2	1	4	7	3	6	8
Group (iv)	patients	44	5	2	1	4	3	7	6	8

TABLE 4. *Normalized factor loadings and eigenvalues for the one factor in ESS item scores for each of four groups of subjects*

ESS item no.	Group (i) patients (n = 150)	Group (ii) students (n = 87)	Group (iii) patients (n = 50)	Group (iv) patients (n = 44)
Normalized factor loadings				
1	0.75	0.57	0.75	0.45
2	0.60	0.48	0.48	0.55
3	0.78	0.68	0.78	0.63
4	0.71	0.62	0.77	0.49
5	0.55	0.44	0.52	0.08
6	0.72	0.26	0.67	0.60
7	0.77	0.71	0.80	0.66
8	0.71	0.41	0.71	0.56
Eigenvalue	3.95	2.32	4.11	2.24

jects; second, by item analysis of those item scores and, third, by factor analysis.

Correlation analysis

A matrix of Spearman correlation coefficients between all 28 pairs of the eight ESS item scores was calculated separately for each of the four groups of subjects. The magnitude of these correlation coefficients was not influenced by the different numbers of subjects in each group, although this did affect their statistical significance.

Twenty of the 28 correlation coefficients for the students in Group (iii) were statistically significant but none represented a close relationship. The mean of all 28 coefficients was 0.27 and the highest was 0.47, between items 3 and 7. Kendall's coefficient of concordance, C, was 0.50. For comparison, the eight correlation coefficients between the students' test and retest scores for the same ESS items had a mean of 0.56 and a maximum of 0.67 (Table 2). The latter eight correlation coefficients were significantly higher than the former 28 (Mann-Whitney *U* test, $p < 0.001$). This means that, in general, a student's SP in a particular situation was significantly more closely related to his SP in the same situation measured 5 months later than it was to his SP in other situations measured at the same time. Thus, the differences between SPs in the same subject were not simply due to unreliable measurements.

Similarly, the mean of the 28 interitem correlation coefficients for the patients in Group (iv), whose spouses also answered the ESS, was 0.42. Twenty-six of the 28 were statistically significant and Kendall's C was 0.55. These patients' own assessments of their sleepiness in each situation were significantly more closely related to their spouses' assessments for the same situations (mean rho = 0.57, Table 2) than they were to their own assessments of SP in other situations (Mann-Whitney *U* test, $p < 0.05$).

The mean of 28 interitem correlation coefficients for the patients in Group (ii) was 0.49, all of which were statistically significant, and Kendall's C was 0.41. The equivalent correlation coefficients for the patients in Group (i) had a mean of 0.21, 13 of which were statistically significant, and Kendall's C was 0.44.

For all four groups of subjects there were significant correlations between many of their SPs in different situations, but they were not close relationships. The SP in one situation usually could not be predicted accurately from measurements of SP in another situation. This is not because the measurements are unreliable. These results have been corroborated in several groups of subjects here.

Item analysis of ESS item scores

The results of item analysis of ESS scores have been reported previously for the patients in Group (ii) and for a group of 104 medical students of which the 87 students in Group (iii) here formed the major part (16). The same analysis was performed here for all four groups of subjects to provide further corroborative evidence about the relationships between ESS item scores.

Cronbach's alpha was 0.74 for the patients in Group (i), 0.88 for Group (ii), 0.75 for the students in Group (iii) and 0.86 for the patients in Group (iv). These results confirm that the ESS has a high level of internal consistency between its items, despite the modest interitem correlation coefficients reported above.

Factor analysis of ESS item scores

Factor analysis of ESS item scores was performed separately for the four groups of subjects. There was only one factor in each (Table 4). The normalized factor loadings differed only slightly between groups, mainly in relation to items 5 and 6. Most factor loadings exceeded 0.5. This analysis confirms that there is a common dimension of variation in ESS item scores that can be interpreted as the subjects' general level of sleepiness across the range of situations included in the ESS.

DISCUSSION

The results raise important issues about the concept and measurement of sleepiness. It is more complicated than previously thought. Measurements of SP are to some extent situation specific, so that a subject's sleepiness in one situation is not necessarily closely related to that in another. This has been demonstrated here in terms of ESS item scores that measure a subject's usual tendency to doze in eight real-life situations. Two new variables have been introduced here—situational SP and average SP. The former is the subject's usual SP in a particular situation, the latter is a more general characteristic derived from a variety of situational SPs measured in daily life. ESS item scores and the SLs measured by the MSLT and MWT are all assessments of different situational SPs. The situations can be described in general terms but not completely, for they depend on the subject's perception of them. It has been demonstrated previously how the SP in a particular situation can be influenced by instructions and motivation (21,22). Nevertheless, within a group of subjects, the different situations in the ESS can be ranked according to their soporific nature. There are consistencies in these rankings between groups but they are not universal.

The situational SPs represented by ESS item scores are relatively constant in a given subject and can be measured with a reliability that is comparable to that for each measurement of SL in the MSLT. Total ESS scores provide an assessment of the subject's average SP that can be measured as reliably as the mean SL in the MSLT (23,24). It is this average SP that is of particular interest when assessing the effects of chronic sleep disorders such as OSA or narcolepsy on patients' sleepiness in daily life.

For the patients in Group (i) here only some of their situational SPs measured by ESS item scores were related significantly to their mean SL in the MSLT. In a multiple regression 40% of the variance in mean SL could be accounted for by ESS item scores, several making significant independent contributions to the regression. Thus, there is some commonality between the situational SPs measured by the two different methods. This is reflected in a correlation coefficient that is statistically significant, but of only moderate magnitude, between total ESS scores and the mean SL in the MSLT ($r = 0.42$, $p < 0.01$).

The correlation between mean SPs in the MSLT and MWT reported by Sangal et al. ($r = 0.41$) (9) is of similar magnitude to that reported here between eight other situations assessed by the ESS. Far from being unusual, it appears that this modest correlation between situational SPs is as should be expected. On average, only about 20% of the variance between different situational SPs represents a global measure of the subject's sleepiness. Some of the remaining variance is situation specific for each subject. This subject × situation specificity makes it possible for some subjects, paradoxically, to fall asleep quicker in the MWT when asked to stay awake than in the MSLT when asked to fall asleep.

For many sleepy people, simply to sit or lie down and relax almost anywhere causes them to fall asleep. The ESS can identify such people with a high sensitivity compared with the MSLT. However, the ESS also identifies subjects who claim to be very sleepy in their daily lives but who do not fall asleep when under scrutiny in the MSLT. To what extent this discordance involves false positive results with the ESS or false negative results with the MSLT requires further investigation. Currently such patients may be diagnosed as having "subjective sleepiness without objective findings", the implication being that they do not really have a disorder of excessive sleepiness affecting their daily lives. The nature of such a disorder remains unexplained.

The present results suggest a need to reevaluate the nature of the sleepiness measured by the MSLT. There is a substantial body of evidence that the MSLT gives reliable, accurate and valid measurements of SP within its own test situation (2). However, the results of Sangal et al. (9) with the MSLT and MWT and the present results with ESS item scores have demonstrated the problems of extrapolating the results from one test situation to another. No one situational SP can be relied upon to represent accurately a subject's average SP in daily life. The total ESS score gives an estimate of this average SP on the basis of eight situational SPs. Currently we do not have a gold standard with which to compare that estimate. All subjective reports such as those in the ESS are open to unintended bias, purposeful falsification or simply the inability of the subject to read and comprehend the questionnaire. These do not appear to have been problems in practice here. However, those possibilities make the ESS of doubtful use in providing legal evidence, for example, in determining fitness to drive a vehicle. In addition, it may be possible for a particular situational SP, such as long-distance driving, to include one component that is subject and situation specific, so that a driver dozed readily at the wheel but was not excessively sleepy in other situations. The MSLT also would not be of much use under such circumstances if its results cannot be transposed reliably to the driver's situation.

The ESS cannot replace the MSLT for measuring short-term changes in sleepiness over periods of hours to days, for example, after taking a sedative drug. The ESS also cannot demonstrate the early onset of rapid eye movement sleep, which is so important in the diagnosis of narcolepsy (2). Nevertheless, the ESS is inexpensive and easy to use with large numbers of subjects and is one method for assessing sleepiness in patients with chronic sleep disorders. All methods for measuring sleepiness should be used with a full understanding of their respective limitations.

REFERENCES

1. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the Multiple Sleep Latency Test (MSLT): a standard measure of sleepiness. *Sleep* 1986;9:519–24.
2. Thorpy MT. The clinical use of the Multiple Sleep Latency Test. *Sleep* 1992;15:268–76.
3. Pressman MR, Fry JM. Relationship of autonomic nervous system activity to daytime sleepiness and prior sleep. *Sleep* 1989; 12:239–45.
4. 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.
5. Mitler MM, Gujavarty KS, Browman CP. Maintenance of Wakefulness Test: a polysomnographic technique for evaluating treatment in patients with excessive somnolence. *Electroencephalogr Clin Neurophysiol* 1982;53:648–61.
6. Van den Hoed J, Kraemer H, Guilleminault C, et al. Disorders of excessive daytime somnolence: polygraphic and clinical data for 100 patients. *Sleep* 1981;4:23–37.
7. Carskadon MA, Dement WC. Daytime sleepiness: quantification of a behavioural state. *Neuro Biobehav Rev* 1987;11:307–17.
8. Poceta JS, Timms RM, Jeong D-U, Ho S-L, Erman MK, Mitler MM. Maintenance of Wakefulness Test in obstructive sleep apnea. *Chest* 1992;101:893–7.
9. Sangal RB, Thomas L, Mitler MM. Maintenance of Wakefulness Test and Multiple Sleep Latency Test: measurement of different abilities in patients with sleep disorders. *Chest* 1992;101:898–902.
10. Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991;14:540–5.
11. Smolley LB, Ivey C, Farkas M, Faucette E, Murphy S. Epworth Sleepiness Scale is useful in monitoring daytime sleepiness. *Sleep Res* 1993;22:389.
12. Roehrs T, Zorick F, Wittig R, Conway W, Roth T. Predictors of objective level of daytime sleepiness in patients with sleep-related breathing disorders. *Chest* 1989;95:1202–6.
13. Poceta JS, Jeong D-U, Ho S-L, Timms RM, Mitler MM. Hypoxemia as a determinant of daytime sleepiness in obstructive sleep apnea. *Sleep Res* 1990;19:269.
14. Johns MW. Daytime sleepiness, snoring and obstructive sleep apnea: the Epworth Sleepiness Scale. *Chest* 1993;103:30–6.
15. Guilleminault C, Partinen M, Quera-Salva MA, Hayes B, Dement WC, Nino-Murcia G. Determinants of daytime sleepiness in obstructive sleep apnea. *Chest* 1988;94:32–7.
16. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep* 1992;15:376–81.
17. Roth T, Roehrs T, Carskadon M, Dement W. Daytime sleepiness and alertness. In: Kryger MH, Roth T, Dement WC, eds. *Principles and practice of sleep medicine.* Philadelphia: Saunders, 1989:14–23.
18. Lavie P. Sleep habits and sleep disturbances in industrial workers in Israel: main findings and some characteristics of workers complaining of excessive daytime sleepiness. *Sleep* 1981;4:147–58.
19. Oswald I. Falling asleep open-eyed during intense rhythmic stimulation. *Brit Med J* 1960;1:1450–5.
20. Hauri PJ, Olmstead EM. Reverse first night effect in insomnia. *Sleep* 1989;12:97–105.
21. Hartse KM, Roth T, Zorick FJ. Daytime sleepiness and daytime wakefulness: the effect of instruction. *Sleep* 1982;5:S107–18.
22. Alexander C, Blagrove M, Horne J. Subject motivation and the Multiple Sleep Latency Test (MSLT). *Sleep Res* 1991;20:403.
23. Zwyghuizen-Doorenbos A, Roehrs T, Schaeffer M, Roth T. Test–retest reliability of the MSLT. *Sleep* 1988;11:562–5.
24. Seidel WF, Dement WC. The Multiple Sleep Latency Test: test–retest reliability. *Sleep Res* 1981;10:284.