# Brief and distinct empirical sleepiness and fatigue scales[☆]

Sally Bailes[a,*], Eva Libman[a,c,e], Marc Baltzan[b,e], Rhonda Amsel[e],
Ron Schondorf[a,e], Catherine S. Fichten[a,d,e]

[a]SMBD-Jewish General Hospital, Montreal, Canada
[b]Mount Sinai Hospital, Montreal, Canada
[c]Concordia University, Montreal, Canada
[d]Dawson College, Montreal, Canada
[e]McGill University, Montreal, Canada

## Abstract

**Objective:** Sleepiness and fatigue are conceptually distinct but pervasively confounded in research, measurement instruments, clinical settings, and everyday spoken language. The purpose of the present study was to construct two scales that represent unconfounded measures of sleepiness and fatigue, using widely used questionnaires. **Method:** Four questionnaires purporting to measure sleepiness [Stanford Sleepiness Scale (SSS); Epworth Sleepiness Scale (ESS)] or fatigue [Fatigue Severity Scale (FSS); Chalder Fatigue Scale (CFS)] were administered, as well as a battery measuring sleep, psychological, and health functioning variables, to three samples: 19 individuals with chronic fatigue syndrome, 14 with narcolepsy, and 11 normal control subjects. **Results:** Analyses revealed two distinct sets of items (six sleepiness and three fatigue items) that were combined into two scales. These newly formed scales are only minimally correlated and represent separate constructs that have reasonably distinctive patterns of association. Findings were replicated and validated in a sample of 128 older individuals complaining of daytime sleepiness and/or fatigue. **Conclusions:** We conclude that (a) it is possible to derive empirically distinct sleepiness and fatigue scales from existing, commonly used self-report instruments, (b) the Empirical Sleepiness Scale is limited to the experience of daytime sleep tendency, while (c) the Empirical Fatigue Scale is associated more broadly with insomnia, psychological maladjustment, and poorer perceived health function. The important clinical implication of the new Empirical Sleepiness and Fatigue Scales is in the ability to identify "sleepiness which is not fatigue," a construct closely related to primary sleep disorders, such as sleep apnea/hypopnea syndrome, for which there is both available and effective treatment.
© 2006 Elsevier Inc. All rights reserved.

*Keywords:* Fatigue; Sleepiness; Empirical; Self-report; Scale; Measure; Questionnaire

## Introduction

Proper differential diagnosis in general medicine and in the mental health domain relies heavily on the accurate distinction between sometimes overlapping symptoms.

Excessive daytime sleepiness and fatigue, highly prevalent in both community and patient populations [1–5], have overlapping features which can lead to imprecise diagnostic formulations and subsequent suboptimal intervention and management decisions.

There is heterogeneity in the definitions of both sleepiness and fatigue [6] as well as in the assessment tools for these constructs [7]. The problem is compounded by the counterintuitive manner in which the constructs sometimes operate. For example, it has long been known that fatigue, rather than sleepiness, is correlated with the experience of insomnia [8]. Even patients with obstructive sleep apnea complain of fatigue, tiredness, and lack of energy at least as

---

* Corresponding author. ICFP-Department of Psychiatry, Jewish General Hospital, 4333 Cote Ste Catherine Road, Montreal, Quebec, Canada H3T 1E4.
*E-mail address:* sally.bailes@mail.mcgill.ca (S. Bailes).

often as they complain of the more expected sleepiness [9]. In addition, scores on self-report measures of daytime sleepiness often correlate only minimally with either self-report [10] or with direct, objective measurement of sleepiness [11].

In a recent study [12], an adjective checklist was constructed describing feeling states related to fatigue and sleepiness. Their five subscales, derived through factor analysis, have high internal consistency and a logical pattern of convergent validity. However, all subscales, most notably sleepiness and fatigue, were highly correlated. In experimental studies, the constructs of fatigue and sleepiness are both separable and additive in their negative effects on performance [13]. In medical practice, sleepiness and fatigue are often equated, leading to inadequate diagnosis and treatment [14]. For example, the specific daytime sleepiness features of sleep apnea are often not recognized, leading to under-referral for further diagnostic evaluation procedures, particularly in the case of women [15]. Daytime fatigue, as distinct from sleepiness, is a concomitant of many physical (e.g., multiple sclerosis, cancer, Parkinson's disease) and psychological (e.g., depression) disorders. A simple, reliable tool to distinguish sleepiness from fatigue made available to health care professionals would assist in the match between symptom identification and appropriate treatment [6].

Because available self-report measures of fatigue and sleepiness are confounded and because the distinction has important consequences for diagnosis, the goals of the present study were (1) to operationalize the terms "sleepiness" and "fatigue" more precisely, (2) to enhance the distinction between them, and (3) to use items from existing measures to prepare empirical-based scales to measure the constructs more accurately. Specifically, we devised and cross-validated "pure" scales of sleepiness and fatigue where the items are empirically derived from existing sleepiness and fatigue measures. We also evaluated scores on these newly developed scales in relation to a range of psychological adjustment, sleep, and perceived physical health instruments in order to develop distinctive sleepiness and fatigue profiles.

## Method

### Overview

The present study was carried out in the context of a larger investigation of chronic fatigue syndrome (CFS) and sleep disorder [15,16]. Here we report on aspects of the procedure and data analysis that pertain to deriving the sleepiness and fatigue scales. To this end, we collected responses to four well-known daytime sleepiness and fatigue questionnaires scales as well as measures of health-related functioning, sleep, and psychological adjustment in three samples: (1) individuals diagnosed with CFS, (2) individuals diagnosed with narcolepsy, and (3) healthy controls with no daytime sleepiness or fatigue complaints. Chronic fatigue syndrome and narcolepsy were selected because the defining symptom of the former is fatigue, and that of the latter is daytime sleepiness. Scores on all sleepiness and fatigue items from the four scales purporting to measure these concepts were correlated. Only items that were not significantly correlated with each other were retained. These comprise the distinct Empirical Sleepiness and Fatigue Scales. Next, an extensive profile of sleepiness and of fatigue was generated by correlating scores on the newly derived empirical scales with scores on the measures of psychological adjustment, sleep, and perceived physical health. Finally, as a replication and validation, these three steps were repeated using a sample of older individuals.

### Participants

#### Development sample

Participants in the groups used to develop the scales were 19 individuals with CFS (all females, mean age=44.7, S.D.=8.5), 14 individuals with narcolepsy (8 females, 6 males, mean age=36.9, S.D.=17.3), and 11 individuals (5 females, 6 males, mean age=40.6, S.D.=9.3) with no diagnosed medical or psychiatric condition, and no complaint of excessive daytime sleepiness or fatigue (control group).

Chronic fatigue syndrome participants were recruited from physician referrals and CFS support groups. For each participant, two independent assessments of CFS were made. Participants arrived with a diagnosis from their own physician. The research team physician confirmed the original CFS diagnosis by using a standardized diagnostic instrument based on the diagnostic criteria of Fukuda et al. [17]. Individuals with narcolepsy were recruited from physician referrals, principally from the Mount Sinai Hospital Sleep Clinic in Montreal. They were diagnosed by information elicited through medical history, overnight polysomnography and daytime multiple sleep latency tests (MSLTs). The usual criteria were evaluated, i.e., presence of sleep attacks, excessive daytime sleepiness, cataplexy, sleep paralysis, hypnagogic hallucinations, sleep disruption, and abnormal timing of REM sleep. Control group participants were recruited from the community through posters, announcements, and personal contacts. Additional details about these groups are available [16].

We used the same pool of participants as described in Fossey et al. [16]. In the present sample, polysomnography evaluation resulted in a diagnosis of sleep disorder in 6 of the 14 narcoleptics [5 apnea/hypopnea syndrome; 1 periodic limb movement disorder (PLMD)], 9 of the 19 chronic fatigue participants (7 apnea/hypopnea syndrome; 2 mild PLMD). Four of the 12 control participants were found to have mild hypopnea symptoms, although they had no complaints. Because of the prevalence of sleep disorders

in our samples, individuals with sleep disorder were not eliminated from the study.

*Validation sample*

An additional 128 older community-based volunteers (59 men and 69 women; mean age=64.8, S.D.=9.4) who participated in a separate sleep-disorders screening study served as a validation sample. These individuals responded to recruiting posters placed in the waiting areas of family practice centers in Montreal hospitals or attended presentations to seniors' groups at community centers (additional details are reported elsewhere [15]). The publicity advertised a research study for individuals suffering from "daytime fatigue or sleepiness or insomnia" and offered a comprehensive evaluation through interview, questionnaire, medical, and polysomnographic assessment. This sample was selected for comparison with the development sample because it provided responses to the same measures by individuals in a different age range and with different clinical characteristics.

All participants gave informed consent. Where physiologically based sleep disorders were diagnosed, the participant was followed and offered treatment by the sleep clinic. In cases where other medical, psychiatric, psychological, or insomnia disorders were diagnosed, appropriate referrals were made.

*Measures: objective measures of fatigue and sleepiness*

*Multiple sleep latency test*

The MSLT is a widely accepted objective behavioral/physiological measure of daytime sleepiness [11]. It consists of giving several nap opportunities during the day and measuring sleep onset latency (i.e., lights out to the first epoch of any sleep stage). In the present study, the absence of sleep was recorded as a latency of 20 min [18]. This measure routinely demonstrates increased sleepiness in normal sleepers who have been sleep-deprived [11] and in individuals with disorders such as narcolepsy and sleep apnea [18].

*Handgrip Fatigue Measure*

This test measures the ability of participants to sustain muscular effort during a period of 30 s. It has been used to provide an objective measure of daytime fatigue [19] and of strength in subjects with CFS [20]. Participants are asked to grip a hand dynamometer (Lafayette Instruments) as tightly as they could and to continue gripping it as tightly as possible for 30 s. Measurements of grip strength (in kilograms) are taken three times during the 30-s period: initially, after 15 s, and after 30 s of sustained grip.

*Measures: sleepiness and fatigue questionnaires—concurrent and retrospective versions*

The instructions for all measures in this section were adapted to have both current (e.g., level of fatigue or sleepiness at this moment) and retrospective (e.g., general level of fatigue in the previous month) versions. The retrospective versions were included in a one-time questionnaire battery administered as part of the sample selection process. The current versions were administered four times throughout 1 day (i.e., at 10 a.m., 12 noon, 2 p.m., 4 p.m.).

*Stanford Sleepiness Scale (SSS)*

This scale, developed by Hoddes et al. [21], is frequently used to assess subjective perceptions of daytime sleepiness. It consists of a seven-point Guttman scaled item ranging from 1 (*feeling active and vital; alert; wide awake*) to 7 (*lost struggle to remain awake*). Respondents select the one option which best describes how sleepy they feel at the moment of testing.

*Epworth Sleepiness Scale (ESS)*

This brief self-administered retrospective questionnaire of the behavioral aspects of sleepiness was constructed by Johns [22] to evaluate self-reports of sleep tendency. Participants rate how likely they are to doze off or fall asleep in eight different situations commonly encountered in daily life on a four-point scale (0=*never doze off*, 3=*high chance of dozing*). Scores are summed and vary from 0 to 24; higher scores indicate greater sleepiness. This measure has high 5-month test–retest reliability in "normals" ($r$=.82), as well as high internal consistency (Cronbach's alpha=.88). Scores are not correlated with SSS scores [23,24].

*Chalder Fatigue Scale (CFM)*

This is an 11-item self-rating scale developed to measure severity of experienced fatigue [25]. The measure has two subscales to evaluate two kinds of fatigue: physical and mental. A total fatigue score is obtained by summing all items. The original version provided four response options: 1="not at all," 2="no more than usual," 3="more than usual," and 4="much more than usual." This was revised for clinical use in our laboratory to use a six-point Likert scale, where 1=*strongly disagree* and 6=*strongly agree*. Subscale scores can be obtained by summing scores on the physical fatigue and on the mental fatigue items. The test has been shown by its authors to have good reliability ($r$=.86 for physical fatigue, and $r$=.85 for mental fatigue) and has high internal consistency as measured by Cronbach's alpha (.89). Validation coefficients for the fatigue scale, using the Revised Clinical Interview Schedule as applied to individuals with CFS, were as follows: sensitivity 75.5 and specificity 74.5. Higher scores indicate greater fatigue.

*Fatigue Severity Scale (FSS)*

Developed by Krupp et al. [26], this nine-item scale assesses "disabling fatigue." The scale's authors report psychometric information that shows that the measure is internally consistent. The single score correlates well with analogue measures and it differentiated controls (mean=2.3, S.D.=0.7) from lupus (mean=4.7, S.D.=1.5) and multiple sclerosis patients (mean=4.8, S.D.=1.3). It could also

predict clinically anticipated changes in fatigue over time. The measure was also shown to be largely independent of depressive symptoms. In addition, it has also been successfully used in insomnia research [27].

## Measures: retrospective questionnaire battery

### Sleep questionnaire

This consisted of an abbreviated version of the retrospective questionnaire used in previous investigations by our team [28,29]. It inquires about typical sleep experiences, including sleep parameters such as sleep onset latency, frequency of nocturnal arousals, total wake time, sleep needed, total sleep time, sleep medication taken, and aspects of sleep lifestyle such as bedtime, time when fell asleep, time of wake up, and time when out of bed. The information provided also allows us to compute sleep efficiency scores (% of bedtime spent asleep) and to obtain ratings of respondents' subjective perceptions of their sleep quality on 10-point Likert-type scales.

Scores based on this measure have acceptable psychometric properties for research use. Test–retest correlations indicate reasonable temporal stability ($r$ values for variables used in this investigation range from .58 to .84), and the pattern of correlations among variables shows logical, highly significant relationships [28]. Convergent validity data indicate significant and high correlations between corresponding scores on the Sleep Questionnaire and on 7 days of self-monitoring on a daily sleep diary [e.g., total sleep time, $r(156)=.82$, $P<.001$; total wake time, $r(146)=.72$, $P<.001$; sleep efficiency, $r(154)=.77$, $P<.001$] [30].

### Structured sleep and medical history

A modified version of the clinical instrument developed by Lacks [31] provides information on inclusion and exclusion criteria, parasomnias, physical disorders, sleep phase disorder, medication use, as well as use of hypnotics and sedatives. Most questions require a yes/no answer with prompts in cases of suspected difficulty. This measure has been successfully used in studies of sleep and aging [28,32,33].

### SF-36 Health Survey

This is a 36-item short form (SF-36) constructed to survey health status in the Medical Outcomes Study [34]. The SF-36 was designed for use in clinical practice and research and assesses eight health domains: (1) limitations in physical activities because of health problems; (2) limitations in social activities because of physical or emotional problems; (3) limitations in usual role activities because of physical health problems; (4) bodily pain; (5) general mental health (psychological distress and well-being); (6) limitations in usual role activities because of emotional problems; (7) vitality (energy and fatigue); and (8) general health perceptions. The measure was constructed either for self-administration or for administration by a trained interviewer. Reliability data were reported from studies carried out on both patient and nonpatient samples [34]. Reliability of the subscales ranged from .64 to .96 in different reference groups, the lowest being for psychiatric patients on the general health subscale. The SF-36 has demonstrable validity in that the subscales were found to correlate with ability to work, utilization of health services, as well as scores on other mental health and quality of life measures. Low scores on all subscales indicate disability due to illness, while high scores indicate better functioning due to relatively good health.

### Beck Depression Inventory (BDI-II)

The 21-item BDI is one of the most frequently used measures of depression [35,36]. As in the original version, on the current revision, too, items are scored on a four-point scale (0–3); scores are summed and produce a range from 0 to 63. Higher scores indicate greater depression. A score over 20 is usually considered indicative of clinical depression, while scores of 13 or less are generally considered nondepressed. Scores from 14 to 19 are generally considered "mildly depressed." Beck et al. [36] report excellent psychometric properties for the scale (internal consistency: $r=.92$; test–retest reliability: $r=.93$). A new feature of the BDI-II revision is that there is a seven-item Primary Care subscale that evaluates the affective and cognitive symptoms of depression independent of fatigue, sleepiness, insomnia, and agitation. Test–retest reliability for this subscale is .82, while internal consistency is .86 [36].

### Brief Symptom Inventory (BSI)

A 53-item self-report psychological symptom inventory, the BSI has subscales for nine symptom dimensions (e.g., depression, anxiety, somatization) and three global indices [37]. It is a brief version of the SCL-90 [38]—a frequently used instrument with acceptable reliability and validity. Validation data indicate correlations from .92 to .98 between the symptom dimensions and global indices of the BSI and the SCL-90 [38]. Lower scores indicate better adjustment.

## Procedure

Following a telephone screening interview, participants in both the development and validation samples underwent the following: a 2-h structured interview and questionnaire session that included the test battery evaluating sleep patterns, health functioning, and psychological adjustment as well as the retrospective versions of the four sleepiness and fatigue measures (i.e., ESS, SSS, FSS, and CFM). Participants in the development sample (i.e., those with CFS, narcolepsy, and controls) also spent 24 h in the sleep laboratory of Mt. Sinai Hospital in Montreal. Following the night of polysomnography, participants retained their EEG montage for the rest of the day. They were administered the Handgrip Fatigue test, the current versions of the four sleepiness and fatigue measures, and the MSLT (20-min nap

opportunity) at four testing times (10:00 a.m., 12:00 noon, 2:00 p.m., 4:00 p.m.).

Individuals with narcolepsy were asked to suspend their CNS stimulant medication (e.g., Ritalin, Modafinil) throughout the laboratory protocol. However, participants who were taking medications on a regular basis where suspending these would cause rebound effects, excessive discomfort, or harm were advised to take them as usual. These included antidepressant medication and benzodiazepines at night. All participants were restricted from caffeine and alcohol consumption throughout the protocol.

The research ethics committees of both the SMBD-Jewish General Hospital and the Mount Sinai Hospital of Montreal approved the research protocol.

## Results

### Relationships among the ESS, SSS, FSS, and CFM measures of sleepiness and fatigue

Table 1A through C shows the correlations among total scores on the four popular sleepiness and fatigue measures across three data sets: Retrospective and current scores in the development sample, and retrospective scores in the validation sample. Of greatest interest is the finding that total scores on sleepiness and fatigue measures correlate highly with each other, thereby demonstrating how the constructs measured are confounded.

Table 1

A. Correlations among existing sleepiness and fatigue measure total scores: development sample current data summed over four trials ($n=45$)

|        | ESS    | SSS    | FSS    | CFM |
|--------|--------|--------|--------|-----|
| ESS    | 1      |        |        |     |
| SSS    | .49**  | 1      |        |     |
| FSS    | .42**  | .83*** | 1      |     |
| CFM    | .53*** | .86*** | .93*** | 1   |

B. Correlations among existing sleepiness and fatigue measure total scores: development sample retrospective data ($n=45$)

|        | ESS    | SSS    | FSS    | CFM |
|--------|--------|--------|--------|-----|
| ESS    | 1      |        |        |     |
| SSS    | .39*   | 1      |        |     |
| FSS    | .16    | .67*** | 1      |     |
| CFM    | .38*   | .68*** | .85*** | 1   |

C. Correlations among existing sleepiness and fatigue measure total scores: validation sample retrospective data ($n=128$)

|        | ESS    | SSS    | FSS    | CFM |
|--------|--------|--------|--------|-----|
| ESS    | 1      |        |        |     |
| SSS    | .29*** | 1      |        |     |
| FSS    | .18*   | .57*** | 1      |     |
| CFM    | .25**  | .63*** | .79*** | 1   |

&ast; $P<.05$.
&ast;&ast; $P<.01$.
&ast;&ast;&ast; $P<.001$.

### Item reduction: correlations among sleepiness and fatigue items

We began by examining Pearson product-moment correlation coefficients among single items on the current version of the four sleepiness and fatigue measures: ESS, FSS, CFM, SSS. Only items not significantly correlated with any item of the opposite construct items were retained. This left only six Sleepiness and three Fatigue items, which then comprised the new Empirical Sleepiness and Fatigue Scales, respectively.

All six Empirical Sleepiness Scale items are from the ESS. Scoring is on a four-point scale (0=*never doze off*, 3=*high chance of dozing*), with a range of 0 to 18: higher scores indicate greater sleepiness. One of the three Empirical Fatigue Scale items is from FSS, and two are from the CFM. Scoring is on a six-point Likert scale (1=*strongly disagree*, 6=*strongly agree*) with a range of 3 to 18; higher scores indicate greater fatigue. Sleepiness and Fatigue Scale items are each summed to yield total scores.

The listing of items comprising the Empirical Fatigue and Empirical Sleepiness Scales as well as item–total and Empirical Fatigue and Sleepiness Scale correlations for both the development and validation samples is presented in Table 2. Cronbach's alpha scores for the Empirical Sleepiness Scale range from .92 to .95, and those for the Empirical Fatigue Scale range from .74 to .86. Correlations between Empirical Fatigue and Sleepiness Scale total scores range from .06 to .33: only one of the three correlations reached significance (at the .03 level). These scores and data presented in Table 2 indicate that the newly developed empirical scales distinguish between self-reports of sleepiness and fatigue, and have good psychometric properties.

### Reliability: test–retest correlations

The development sample completed the current version of the Empirical Fatigue and Sleepiness Scales four times: 10 a.m., 12 noon, 2 p.m., and 4 p.m. We correlated scores on each of the nine sleepiness and fatigue items that comprise the Empirical Fatigue and Sleepiness Scales administered at 10 a.m. with scores obtained at 2 p.m. We also ran correlation analyses between scores gathered at the 12 noon and 4 p.m. test times. The resulting 18 Pearson product-moment correlation coefficients ranged from .50 to .91. All reached significance at the .05 level or better. Similarly, test–retest correlations between total Empirical Sleepiness Scale scores were .69 and .88; coefficients for total Empirical Fatigue Scale scores were .87 and .91. All total score correlations were significant at the .001 level.

### Group differences: Empirical Sleepiness and Fatigue Scale scores

To evaluate group differences, a multivariate analysis of variance (MANOVA) test was carried out comparing the

Table 2
Empirical Sleepiness and Fatigue Scales: item/total correlations for both samples at all testing times

| | Development sample, current ($n$=45) | | Development sample ($n$=45), retrospective | | Validation sample ($n$=128), retrospective | |
|---|---|---|---|---|---|---|
| | Empirical Sleepiness Scale | Empirical Fatigue Scale | Empirical Sleepiness Scale | Empirical Fatigue Scale | Empirical Sleepiness Scale | Empirical Fatigue Scale |
| *Empirical Sleepiness Scale items*[a] | | | | | | |
| How likely are you to doze off or fall asleep in the following situations, in contrast to just feeling tired? | | | | | | |
| Sitting and reading | .91* | .20 | .82* | .18 | .79* | .06 |
| Watching TV | .95* | .22 | .88* | .14 | .79* | .08 |
| Sitting inactive in a public place (e.g., theatre, meeting) | .90* | .21 | .91* | .17 | .84* | .10 |
| As a passenger in a car for an hour when circumstances permit | .88* | .19 | .85* | .17 | .73* | .13 |
| Sitting and talking to someone | .80* | .19 | .78* | .22 | .56* | .14 |
| Sitting quietly after lunch without alcohol | .92* | .12 | .87* | .03 | .69* | .01 |
| *Empirical Fatigue Scale items*[b] | | | | | | |
| Exercise brings on my fatigue[c] | .13 | .91* | .13 | .86* | .01 | .67* |
| I start things without difficulty but get weak as I go on[d] | .16 | .84* | .13 | .88* | .01 | .75* |
| I lack energy[d] | .28 | .92* | .21 | .86* | .04 | .78* |

[a] All items from the Epworth Sleepiness Scale [22]. Scoring is on a 4-point scale (0 never doze off, 3 high chance of dozing), with a range of 0 to 18: higher scores indicate greater sleepiness.

[b] Scoring is on a 6-point Likert scale (1=strongly disagree, 6=strongly agree) with a range of 3 to 18; higher scores indicating greater fatigue.

[c] Item from the FSS [26].

[d] Items from the CFS [25].

\* $P$<.01.

three groups in the development sample (i.e., CFS, narcolepsy, control) on the Empirical Sleepiness and Fatigue Scales, the MSLT and Handgrip scores. The MANOVA was significant, $F(8,60)$=10.68, $P$<.001. ANOVA and Tukey HSD post hoc tests reported in Table 3 revealed that on the Empirical Sleepiness Scale, both the Narcolepsy and CFS groups had significantly higher sleepiness scores than the Control group. On the Empirical Fatigue Scale, all three groups differed significantly, with the CFS group having the highest fatigue scores, followed by the Narcolepsy group, followed by the Control group. The three groups differed significantly from each other on both the MSLT and Handgrip tests: The Narcolepsy group had the shortest sleep latencies, and the CFS group had the least grip strength.

*Correlations with other measures*

For the development sample, data from the objective test (i.e., MSLT, Handgrip) were averaged across the four testing times and correlated with the total scores on the new Empirical Scales. The Empirical Sleepiness Scale was not significantly correlated with either the MSLT or the Handgrip scores. The correlation between scores on the Empirical Fatigue Scale and the Handgrip test was significant, $r$=−.33, $P$=.027, suggesting that increased self-rated fatigue was associated with lower grip strength. Score on the Empirical Fatigue Scale was also negatively correlated with score on the MSLT, $r$=−.40, $P$=.014, suggesting discrimination between self-rated fatigue and objective sleep propensity. Although these correlations are

Table 3
Chronic fatigue syndrome, narcolepsy, and control group: means, standard deviations, and test results for Empirical Sleepiness Scale, Empirical Fatigue Scale, MSLT, and Handgrip

| | CFS[a] ($n$=19) | Narcolepsy[a] ($n$=14) | Control[a] ($n$=12) | Test results | | | |
|---|---|---|---|---|---|---|---|
| | | | | $F$ | $df$ | $P$ | Post hoc |
| Empirical Sleepiness Scale | 7.4 (6.1) | 8.2 (4.0) | 3.1 (3.1) | 3.98 | 2,41 | .026 | N, CFS>C |
| Empirical Fatigue Scale | 18.0 (4.8) | 11.2 (5.3) | 8.8 (4.0) | 16.24 | 2,42 | .000 | CFS>N, C |
| MSLT (minutes to sleep onset) | 16.0 (5.3) | 5.1 (6.0) | 11.3 (4.8) | 12.87 | 2,34 | .000 | N<C<CFS |
| Handgrip strength 30 (kg) | 8.8 (4.3) | 14.3 (5.4) | 20.0 (8.1) | 13.64 | 2,42 | .000 | C>N>CFS |

CFS, chronic fatigue syndrome; N, Narcolepsy; C, control.

[a] Values are mean (S.D.).

Table 4
Correlations between retrospective Empirical Sleepiness and Fatigue Scale totals with retrospective test battery scores evaluating sleep, psychological, and health functioning in the development and validation samples

| Retrospective questionnaire battery | Empirical Sleepiness Scale | | Empirical Fatigue Scale | |
|---|---|---|---|---|
| | Development sample ($n=45$) | Validation sample ($n=128$) | Development sample ($n=45$) | Validation sample ($n=128$) |
| Do you wake up in the middle of the night feeling unable to breathe? | .48** | .08 | .11 | .13 |
| Have you noticed that parts of your body jerk at night? | .47** | .04 | .08 | .01 |
| Do you have difficulty staying awake during the day when you really want to be awake? | .58** | .37*** | .38 | .08 |
| Do you have difficulty staying awake at awkward times, e.g., while you are driving, at a table with friends, while at work, etc.? | .59*** | .33*** | .19 | .05 |
| How many days per week do you usually nap during the day? | .41** | .31*** | .29 | .22 |
| Generally, how sleepy do you feel during the day? | .51*** | .34*** | .49*** | .31** |
| Generally, how difficult is it to concentrate on what you have to do? | .35 | .22 | .57*** | .38*** |
| Do you feel exhausted during the day? | .27 | .10 | .54*** | .48*** |
| Do you have any illnesses? | .19 | .09 | .41** | .05 |
| Do you have insomnia? | .12 | −.07 | .48*** | .10 |
| I do not feel refreshed when I get up in the morning | −.27 | −.04 | .40** | .26** |
| Generally, what is the quality of your sleep? | −.32 | −.15 | −.59*** | −.13 |
| Generally, how satisfied are you with your sleep? | −.35 | −.15 | −.66*** | −.05 |
| How refreshed do you usually feel in the morning? | −.21 | −.01 | −.56*** | −.34*** |
| Generally, how tired do you feel during the day? | .27 | .06 | .72*** | .47*** |
| Beck Depression Inventory total | .22 | .03 | .42** | .41*** |
| SF-36 Physical functioning | −.04 | −.06 | −.77*** | −.38*** |
| SF-36 Role physical | −.17 | −.03 | −.73*** | −.43*** |
| SF-36 Bodily pain | −.17 | −.04 | −.64*** | −.22 |
| SF-36 General health | .02 | .00 | −.75*** | −.37*** |
| SF-36 Vitality | −.17 | −.04 | −.70*** | −.58*** |
| SF-36 Social functioning | −.07 | .01 | −.62*** | −.35*** |
| SF-36 Role emotional | −.25 | −.08 | −.02 | −.22 |
| SF-36 Mental health | −.16 | −.10 | −.12 | −.25** |

*P<.05, not shown.
To offset the effect of multiple correlations, the significance level of coefficients is indicated only when they reach a minimum .01 criterion. Italicized sections highlight correlates of Empirical Sleepiness and Empirical Fatigue Scale total scores.
   ** P<.01.
   *** P<.001.

significant, they are nevertheless small and difficult to interpret. They are included here for the reader's interest.

Correlations between total scores on the retrospective version of the two empirical scales and scores on the retrospective test battery were calculated for both the development and validation samples. Results in Table 4 show the correlation coefficients; to offset the effect of multiple correlations, the significance level of coefficients is indicated only when they reach the .01 criterion.

It can be seen in Table 4 that total scores for the Empirical Sleepiness Scale were generally related to experienced daytime sleepiness, frequency of daytime naps, and breathing difficulty at night. In contrast, total Empirical Fatigue Scale scores were associated with a range of variables, including daytime fatigue, perceived quality of daytime functioning, the insomnia complaint, depression, and aspects of perceived health quality. It is noteworthy that score on the item "Generally, how sleepy do you feel during the day?" was significantly correlated with both Empirical Fatigue and Empirical Sleepiness Scale totals.

## Discussion

### Existing measures of sleepiness and fatigue

When scores on two well-known and frequently used measures of daytime sleepiness and two well-known and frequently used measures of fatigue were examined, we found generally high and significant correlations between the two types of measures. In fact, in one case the sleepiness measure (SSS [21]) correlated more highly with the two fatigue measures than with the other sleepiness measure (ESS [22]). This indicates that these instruments seriously confound the concepts of sleepiness and fatigue as well as their measurement.

### The new Empirical Sleepiness and Empirical Fatigue Scales

When we derived individual sleepiness and fatigue items that were not related to the opposite construct, we identified six sleepiness items, derived exclusively from the ESS [22], and three fatigue items out of the 20 comprising both the

FSS [26] and the CFS [25]. Sleepiness items derived in this way are related exclusively to the respondent's chance of dozing during a variety of daytime situations. Fatigue items are related to weakness and tiredness resulting from physical exercise and other daytime activities as well as to a general perceived lack of energy.

Our analyses indicate good test–retest reliability and internal consistency for both empirical scales, although the test–retest interval was only 4 h, and this result needs to be replicated. Scores on the separate items within the two empirical scales do not correlate with total scores on the scale measuring the opposite construct. The total scores of the two empirical scales are not correlated significantly. These results are consistent for samples of individuals who differ widely on age and health status. Furthermore, individuals with Narcolepsy and CFS differed from Controls on both newly developed Empirical Scales: CFS subjects had higher Empirical Fatigue Scale scores than those with narcolepsy, indicating good discrimination for the scales.

Our statistical analyses corroborated Pigeon et al.'s [6] observation of heterogeneity in current definitions of sleepiness and fatigue. Items in the original sleepiness and fatigue measures contain descriptors such as "tired," "drowsy," and "poor concentration." These blurred the distinction between the constructs of sleepiness and fatigue, and were eliminated from the new empirical scales.

*What is measured by the new Empirical Sleepiness and Fatigue Scales?*

There was a logical pattern of correlations between the new Empirical Sleepiness and Fatigue Scales with the other behavioral and psychophysiological measures in our two clinical samples. The narcolepsy group had the highest Empirical Sleepiness Scale scores and the shortest latencies on the MSLT, while the CFS group had the highest Empirical Fatigue Scale scores and the weakest handgrip strength. The constellation of correlations between "pure" sleepiness, as measured by the newly developed Empirical Sleepiness Scale, and other cognitive-affective and behavioral measures clearly reflected physical experiences which disrupt sleep at night (e.g., feeling unable to breathe, involuntary movements), and experienced daytime drowsiness (perceived impaired alertness, inclination to doze inappropriately during activities, and taking naps).

The Empirical Sleepiness Scale did not correlate with performance on the MSLT, an assumed objective sleepiness measure. It is possible that the high percentage of people with sleep disorders in our samples may have obscured this correlation. Also, it is well known that individuals suffering from nocturnal insomnia manifest the same problem when instructed to fall asleep in the daytime [39]. Alternately, the lack of correlation may simply be another example of a well-documented finding in the literature, i.e., that sleep propensity, as measured by the MSLT, is a different construct from subjective sleepiness/alertness, as measured by self-report [40].

"Pure" fatigue, as measured by the newly developed Empirical Fatigue Scale, is anything but pure, because it is associated with many aspects of functioning. It clearly reflects experienced nonrestorative sleep as well as daytime exhaustion. It also correlates significantly with objective fatigability as measured by the handgrip test. However, fatigue scores were also highly and significantly correlated with perceived impairment of psychological and physical health, ability to function generally, and quality of life.

Only the Empirical Fatigue Scale was found to be significantly related to the insomnia complaint and its manifestations, while the Empirical Sleepiness Scale was not related to insomnia variables. Although this may seem counterintuitive, this pattern of findings underlines previous indications in the literature that insomnia is not synonymous with sleep deprivation and that people with insomnia are more likely to be tired than sleepy [8,41].

Items dealing with unwanted daytime sleep episodes were not significantly associated with scores on the Empirical Fatigue Scale. However, items reflecting a sleepiness feeling state (e.g., "How sleepy do you feel during the day?") are correlated with scores on both the Empirical Sleepiness and Fatigue Scales. This suggests that only self-reported daytime sleepiness that is related to a daytime sleep-related behavior discriminates between sleepiness and fatigue (e.g., tendency to fall asleep in inappropriate places).

Both the sleepiness and fatigue constructs may be more complex than represented by these "pure" empirical scales: the correlates of "pure" sleepiness and "pure" fatigue need further investigation. In particular, further investigation in experimental studies of sleep deprivation should be carried out, and the scales should be administered in studies involving primary sleep disorders such as sleep apnea/hypopnea syndrome and restless legs/PLMD as well those involving shift work, sleep phase disturbance, and jet lag. Application of these subscales to more diverse groups, including clinical and normal samples of varying ages, would help establish norms and cutoff scores for clinically significant symptoms. This would have obvious research and clinical utility.

In summary, our new Empirical Sleepiness and Empirical Fatigue Scales consist of two "pure" daytime measures. The Empirical Sleepiness Scale appears to be specifically relevant to the likelihood of falling asleep during daytime activities. The Empirical Fatigue Scale appears to be related to a wider range of variables including perceived poor physical and psychological functioning as well as physical tiredness. At present, the main usefulness of the Empirical Sleepiness and Empirical Fatigue Scales is in their ability to identify "sleepiness which is not fatigue," a condition that seems likely in populations suffering from primary sleep disorders such as sleep apnea/hypopnea syndrome, for which there is both available and effective treatment.

# References

[1] Addington AM, Gallo JJ, Ford DE, Eaton WW. Epidemiology of unexplained fatigue and major depression in the community: the Baltimore ECA follow-up, 1981–1994. Psychol Med 2001;31(6): 1037–44.

[2] Cathebras PJ, Robbins JM, Kirmayer LJ, Hayton BC. Fatigue in primary care: prevalence, psychiatric comorbidity, illness behavior, and outcome. J Gen Intern Med 1993;7(3):276–86.

[3] Loge JH, Ekeberg O, Stein K. Fatigue in the general Norwegian population: normative data and associations. J Psychosom Res 1998; 45:53–65.

[4] National Sleep Foundation. Omnibus Sleep in America Poll. Washington (DC): National Sleep Foundation, 2001.

[5] Pawlikowski T, Chalder T, Hirsch SR, Wallace P, Wright DJM, Wessely SC. Population based study of fatigue and psychological distress. BMJ 1994;308:763–6.

[6] Pigeon WR, Sateia MJ, Ferguson RJ. Distinguishing between excessive daytime sleepiness and fatigue: toward improved detection and treatment. J Psychosom Res 2003;54:61–9.

[7] Guilleminault C, Brooks SN. Excessive daytime sleepiness. A challenge for the practicing neurologist. Brain 2001;124:1482–91.

[8] Chambers MJ, Keller B. Alert insomniacs: are they really sleep deprived? Clin Psychol Rev 1993;13:649–66.

[9] Chervin RD. Sleepiness, fatigue, tiredness, and lack of energy in obstructive sleep apnea. Chest 2000;118(2):372–9.

[10] Alapin I, Fichten CS, Libman E, Creti L, Bailes S, Wright J. How is good and poor sleep in older adults and college students related to daytime sleepiness, fatigue and ability to concentrate? J Psychosom Res 2001;49(5):381–90.

[11] Carskadon MA. Measuring daytime sleepiness. In: Kryger MH, Roth T, Dement WC, editors. Principle and practice of sleep medicine. Philadelphia: WB Saunders, 1989. pp. 684–8.

[12] Shapiro CM, Flanigan M, Fleming JAE, Morehouse R, Moscovitch J, Plamondon J, Reinish L, Devins GM. Development of an adjective checklist to measure five FACES of fatigue and sleepiness: data from a national survey of insomniacs. J Psychosom Res 2002; 52:467–73.

[13] Philip P, Sagaspe P, Taillard J, Moore N, Guilleminault C, Sanchez-Ortuno M, Akerstedt T, Bioulac B. Fatigue, sleep restriction, and performance in automobile drivers: a controlled study in a natural environment. Sleep 2003;26(3):277–80.

[14] Lichstein KL, Means MK, Noe SL, Aguillard RN. Fatigue and sleep disorders. Behav Res Ther 1997;35:733–40.

[15] Bailes S, Baltzan M, Alapin I, Fichten CS, Libman E. Diagnostic indicators of sleep apnea in older women and men: a prospective study [in press].

[16] Fossey ME, Libman E, Bailes S, Baltzan M, Schondorf R, Amsel R, Fichten CS. Sleep quality and psychological adjustment in chronic fatigue syndrome. J Behav Med 2004;27(6):581–605.

[17] Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff AL. The chronic fatigue syndrome: a comprehensive approach to the definition and study. Ann Intern Med 1994;121:953–9.

[18] Thorpy MJ, Westbrook P, Ferber R, Fredrickson P, Mahowald M, Perez-Guerra F, Reite M, Smith P. The clinical use of the multiple sleep latency test. Sleep 1992;15:268–76.

[19] Rantanen T, Masaki K, Foley D, Izmirlian G, White L, Guralnik JM. Grip strength changes over 27 yr in Japanese-American men. J Appl Phys 1998;85:2047–53.

[20] Blackwood SK, MacHale SM, Power MJ, Goodwin GM, Lawrie SM. Effects of exercise on cognitive and motor function in chronic fatigue syndrome and depression. J Neurol Neurosurg Psychiatry 1998;65: 541–6.

[21] Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. Psychophysiology 1973;10:431–7.

[22] Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. Sleep 1991;14(6):540–5.

[23] Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep 1994;15(4):376–81.

[24] Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. Sleep 1994;17(8):703–10.

[25] Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, Wallace EP. Development of a fatigue scale. J Psychosom Res 1993;37(2):147–53.

[26] Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The Fatigue Severity Scale. Arch Neurol 1989;46:1121–3.

[27] Lichstein KL, Wilson NM, Noe SL, Aguillard RN, Bellur SN. Daytime sleepiness in insomnia: behavioral, biological and subjective indices. Sleep 1994;17:693–702.

[28] Fichten CS, Creti L, Amsel R, Brender W, Weinstein N, Libman E. Poor sleepers who do not complain of insomnia: myths and realities about psychological and lifestyle characteristics of older good and poor sleepers. J Behav Med 1995;18(2):189–223.

[29] Fichten CS, Libman E, Creti L, Amsel R, Tagalakis V, Brender W. Thoughts during awake times in older good and poor sleepers: the self-statement test: 60+. Cogn Ther Res 1998;22(1):1–20.

[30] Libman E, Fichten CS, Bailes S, Amsel R. Sleep questionnaire vs sleep diary: which measure is better? Int J Rehabil Health 2000; 5(3):205–9.

[31] Lacks P. Daily sleep diary. In: Lacks P, editor. Behavioral treatment for persistent insomnia. New York: Pergamon Press, 1987. pp. 70–3.

[32] Libman E, Creti L, Amsel R, Brender W, Fichten CS. What do older good and poor sleepers do during periods of nocturnal wakefulness? The Sleep Behaviors Scale: 60+. Psychol Aging 1997; 12(1):170–82.

[33] Libman E, Creti L, Levy RD, Brender W, Fichten CS. A comparison of reported and recorded sleep in older poor sleepers. J Clin Geropsychol 1997;3(3):199–211.

[34] Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Lincoln (RI): QualityMetric Incorporated, 2000.

[35] Beck A, Steer R, Brown G. BDI-II: Beck Depression Inventory manual—second edition. San Antonio: The Psychological Corporation, Harcourt Brace & Company, 1996.

[36] Beck A, Guth D, Steer R, Ball R. Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. Behav Res Ther 1997;35:785–91.

[37] Derogatis LR, Rickels K, Rock AF. The SCL-90 and the MMPI: a step in the validation of a new self-report scale. Br J Psychiatry 1976;128:280–9.

[38] Derogatis LR. The psychopathology rating scale: a brief description. Unpublished manuscript, 1977.

[39] Bonnet MH, Arand DL. Insomnia, metabolic rate and sleep restoration. J Intern Med 2003;254:23–31.

[40] Monk TH. Circadian rhythms in subjective activation, mood, and performance efficiency. In: Kryger MH, Roth T, Dement WC, editors. Principle and practice of sleep medicine. Philadelphia: WB Saunders, 1989. pp. 163–72.

[41] Fichten CS, Libman E, Bailes S, Alapin I. Characteristics of older adults with insomnia. In: Lichstein KL, Morin CL, editors. Treatment of late life insomnia. London: Sage, 2000. pp. 37–80.