Sleep, 17(3):236–241

Relations Between Performance and Subjective Ratings of Sleepiness During a Night Awake

Mats Gillberg, Göran Kecklund and Torbjörn Åkerstedt

Karolinska Institute and National Institute for Psychosocial Factors and Health, Stockholm, Sweden

Summary: The present study validated the nine-point Karolinska Sleepiness Scale (KSS) and the new Accumulated Time with Sleepiness (ATS) scale against performance of laboratory tasks. The ATS scale was designed as a method for integrating subjective sleepiness over longer time periods. The subjects were asked if certain symptoms of sleepiness had occurred and, if so, for how long. Six subjects participated twice. Each time they were kept awake during the night (except for a short nap occurring during one of the nights in a counterbalanced order) and were tested at 2200, 0200, 0400 and 0600 hours. The tests included a 10-minute rest period, a 28-minute visual vigilance task and an 11-minute single reaction time task. KSS and visual analogue scale (VAS) ratings were given before each test, and ATS ratings were given after. Performance deteriorated clearly, and all three rating scales reflected increased sleepiness with time of night. Scores on the KSS and VAS showed high correlations with performance tasks (mean intraindividual correlations were between 0.49 and 0.71). Performance correlated even higher with the ATS ratings ($r = 0.73$–0.79). Intercorrelations between rating scales were also high ($r = 0.65$–0.86). It was concluded that there were strong relations between ratings of sleepiness and performance, that the ATS rating scale was at least as good as the other scales and that the ratings were affected by type of task. **Key Words:** Sleepiness—Self-rating scales—Performance—Sleep loss.

Subjective ratings are often the only possible methods for assessing sleepiness in field studies. It is important to stress that subjective methods, apart from being practical in field studies, also evaluate an important aspect of sleepiness. In work situations with more or less passive supervision, such as driving tasks or control room tasks, the individual has no continuous feedback on the quality of performance. The subjective signals of sleepiness are the only information on which the individual bases his decisions about when to discontinue work to avoid mistakes or accidents. Consequently, it is important to know with what precision performance can be predicted by subjective ratings of sleepiness. In our own research we have used subjective methods for evaluations with train drivers (1), truck drivers (2) and three-shift workers (3). To measure absolute levels of sleepiness in contrast to the relative levels of, for example, visual analogue scales (VAS), we recently introduced the nine-point Karolinska Sleepiness Scale (KSS; 4). KSS was found to be strongly related to electroencephalogram (EEG) and electrooculogram signs of sleepiness (4). However, the KSS has not yet been validated against performance. To do so was one of the aims of this study.

Like the Stanford Sleepiness Scale (SSS), KSS requires the subject to integrate and translate a number of sensations to a continuum that is fairly abstract in spite of the verbal descriptions. Ratings obtained with these methods may be influenced by how different individuals cope with this procedure. A less abstract way of rating would be to report directly the duration of the experience of these sensations or symptoms of sleepiness. We have recently developed a rating procedure along these lines, the Accumulated Time with Sleepiness Scale (ATS) (2). The subject is required to report whether a number of symptoms (e.g. heavy eyelids) have appeared during a defined period and, if so, how long these symptoms lasted. Subjects may experience different symptoms of sleepiness, and different situations or tasks may provoke different symptoms. The suggested procedure accommodates such differences and might use this variance to increase psychometric sensitivity. Furthermore, when using SSS, KSS or VAS, subjects are often asked to rate how they feel "right now," how they felt during the last 10 or 15 minutes, or perhaps how they have felt since the last time they rated. In a field situation it might only be

Accepted for publication December 1993.
Address correspondence and reprint requests to Mats Gillberg, Department of Clinical Neuroscience, Section for Stress Research, Karolinska Institute, Box 220, S-171 77 Stockholm, Sweden.

possible to give self-ratings during pauses at work. The ATS was designed to give an integrated rating representing sleepiness during specific tasks or situations longer than 10–15 minutes.

The relation between performance and subjective sleepiness has been investigated in a number of studies. Glenville and Broughton (5) found significant correlations between SSS and performance during the day after a night without sleep. In a study of the effects of cumulative partial sleep deprivation, Herscovitch and Broughton (6) found that SSS was sensitive to the partial deprivation procedure but found no significant correlation between performance and SSS. Johnson et al. (7,8) found that subjective ratings were only marginally related to performance and not at all related to sleep latency on the multiple sleep latency test or lapses during a tapping task. Hence, the results from these studies are somewhat conflicting. However, to achieve high correlations between, for example, subjective ratings and performance, factors presumably affecting sleepiness must be manipulated to a sufficient degree. That is, the duration of prior waking, prior sleep duration and circadian phase must be chosen to ensure a large variation in sleepiness, in order to ensure adequate sampling across the entire rating scale. In the latter three studies this requirement might not have been met. Hence, we decided to follow subjects during a waking night when a rapid increase in sleepiness might be expected due to the combined effects of time awake and circadian phase (9,10).

One purpose of the present experiment was to study the relations between performance and KSS and, for comparison, VAS. Another purpose was to validate the ATS against performance and, furthermore, to study whether response to the ATS reflected the effect of task on alertness. A final purpose was to study interrelations between responses to the three self-rating instruments.

METHODS

Six subjects (two female and four male, age range 25–45 years) participated twice [night 1 (N1) and night 2 (N2)]. There was at least 1 week between the 2 nights and the subjects were studied in pairs. On each occasion the subjects came to the laboratory at 2100 hours and were kept awake until 0700 hours, except for a 1-hour nap, which started at 2300 hours during one of the nights (scheduled in a counter-balanced order between N1 and N2). The naps were included in the design for other experimental purposes than those at focus in this report. Tests and ratings were presented in identical blocks (Fig. 1). The sequence was as follows: KSS and VAS ratings followed by the Karolinska Alertness Test (KAT), constituting a 10-minute period

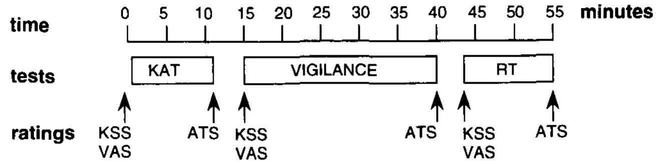

FIG. 1. Sequence of events within one test block. The tests were KAT (Karolinska Alertness Test), visual vigilance task and RT (single reaction time task). The rating scales were KSS (Karolinska Sleepiness Scale), VAS (visual analogue scale) and ATS (Accumulated Time with Sleepiness Scale). Identical blocks were presented at 2200, 0200, 0400 and 0600 hours each of the 2 nights.

of supine rest in bed, with eyes open for 5 minutes followed by another 5 minutes with closed eyes and the instruction to stay awake (the EEG data from the KAT are not reported here). Immediately after completion of the KAT, ratings of symptoms of sleepiness during this test period were given (ATS; see below). The subject then moved to a sitting position in front of a computer and completed the KSS and VAS. A 28-minute visual vigilance task followed that ended with a new ATS rating. After approximately 5 minutes, immediately before starting the 11-minute visual single reaction time test, KSS and VAS ratings were obtained. A final ATS rating was completed after the reaction time test. The entire block lasted approximately 55 minutes. Four blocks were presented at 2200, 0200, 0400 and 0600 hours each of the 2 nights. The subjects participated two at a time in separate sound-attenuated rooms containing a bed, a chair and a computer. The subjects spent their "free time" in their respective rooms, reading or working. Coffee or caffeinated beverages were not allowed.

The vigilance task was presented on a computer screen and consisted of a yellow dot, 5 mm in diameter, jumping sequentially in fixed positions around the periphery of a circle (70 mm in diameter) against a dark background. There were 14 equally distributed positions around the periphery. The dot was visible for 0.45 second and there was an interval of 0.45 second between successive presentations. The signal event was when the dot skipped one position and the interval was increased by 0.3 second. There was a total of 60 signals during the 28 minutes. They appeared in a pseudorandom order with 2–60 seconds between signals.

The reaction time task was a version of the Lisper and Kjellberg (11) task. The signal was presented on a computer screen. Subjects were instructed to press the space bar on the computer keyboard as soon as the word "press" appeared on the screen. The task lasted 11 minutes during which 176 signals were presented. The interstimulus interval varied randomly between 2 and 7 seconds.

The KSS is a nine-point scale ranging from 1 ("very alert") to 9 ("very sleepy, fighting sleep, an effort to keep awake") with verbal descriptions of every second

TABLE 1. F- *and p-values from two-way ANOVAs (performance tasks) and three-way ANOVAs (self-ratings). Where applicable, epsilon values are given.*

Variables	Between nights *F*	Between nights p	Across time of night *F*	Across time of night p	Across time of night Epsilon	Interaction *F*	Interaction p	Interaction Epsilon			
Vigilance (% hits)	0.00	ns	11.22	<0.01	0.51	0.23	ns	0.80			
Mean RT (msec)	1.32	ns	11.06	<0.01	0.42	1.11	ns	0.76			
						Across tests			p and range of *F* for interactions[a]		Range of epsilon[a]
KSS	0.23	ns	24.05	<0.01	0.43	8.10	<0.025	0.66	ns	0.25–0.90	0.33–0.95
VAS	0.05	ns	17.77	<0.01	0.39	11.84	<0.025	0.61	ns	0.89–1.93	0.32–0.78
ATS	0.16	ns	19.27	<0.001	0.51	6.08	<0.025	0.80	ns	0.24–0.70	0.19–0.62

[a] The ranges of *F*- and epsilon values for the possible interactions (night × time of night, night × test, time of night × test, night × time of night × test).

point (4). Apart from the KSS, VAS was used to assess subjective sleepiness. The latter scale was a 100-mm-long line ranging from "very alert" to "very sleepy". KSS and VAS ratings were obtained before the tests because the main interest for these scales was their ability to predict performance. To record them also immediately after each test was not considered meaningful, because it would have been difficult to motivate the subjects to rate their sleepiness so frequently and within such short intervals.

The version of the ATS used in the present study consisted of eight items. The subjects were asked "did you experience any of the following symptoms during the test and, if so, for how long?". The response categories were did not occur, occurred a few times, occurred during 25% of the time, occurred during 50% of the time, occurred 75% of the time and occurred most of the time. The ratings were scored from 0 to 5. The symptoms were heavy eyelids, sand in your eyes, difficulties in focusing your eyes, irresistible sleepiness, difficulties in keeping your eyes open, difficulty focusing attention, difficulty concentrating and periods when you were fighting sleep. ATS ratings were recorded immediately after each test, because it was conceived a method for rating sleepiness experienced during a defined task or situation.

Statistics. The primary purpose of the present analysis was to demonstrate concomitant variation in the variable pairs studied. For that reason variation due to differences in levels between subjects was removed by standardizing data on an individual basis. Each subject yielded eight data points (2 nights, four recordings each night). Standard scores were calculated across these eight data points for each subject. The standardized data from the six subjects were pooled. Hence, the correlation analyses were performed with 6 × 8 = 48 data points for each variable. The use of standardized and pooled data meant that some of the assumptions required for statistical inference may not have been met. However, assumptions of normality will be approximately true if the number of data points is reasonably large and if there are no extreme outliers. This can be demonstrated using limit theorems similar to those used for permutation tests (12). Skewness was calculated for all normalized variables (with n = 48 data points). The skewness values ranged between −0.98 and 0.25, indicating no serious deviations from normality for any of the variables. Furthermore, residual analyses were performed for the linear regressions corresponding to the coefficients of correlation

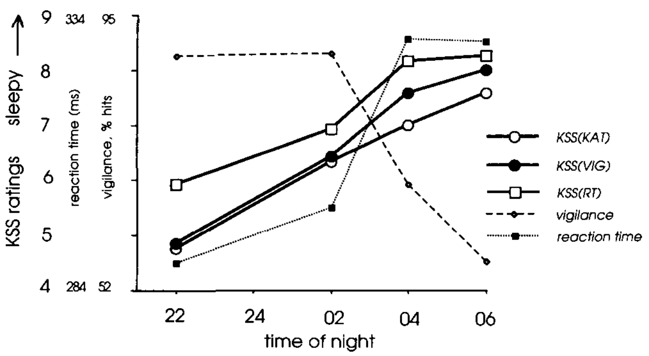

FIG. 2. KSS self-ratings preceding the KAT, vigilance test and reaction time test across time of night. Performance data from the vigilance and reaction time tests are also included. Data from both nights are averaged; means of six subjects.

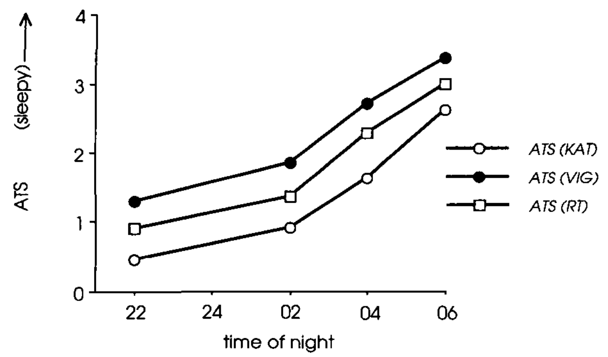

FIG. 3. ATS self-ratings (average of the eight symptoms) following the KAT, vigilance test and reaction time test across time of night. Data from both nights are averaged; means of six subjects. 2 = symptoms during 25% of the test, 5 = symptoms occurred most of the time.

TABLE 2. *Relations between self-ratings and performance. KSS and VAS ratings were given before the tests and ATS ratings were given after the tests. Coefficients of correlations are based on standardized and pooled data from six subjects and eight measurements (n = 48). p-values with df = 41 (see text)*

	Vigilance test	Reaction time test
KSS	−0.62 ($p < 0.0001$)	0.71 ($p < 0.0001$)
VAS	−0.49 ($p < 0.001$)	0.56 ($p < 0.0001$)
ATS	−0.79 ($p < 0.0001$)	0.73 ($p < 0.0001$)

presented in Tables 2 and 3. The range of studentized residuals was between −3.02 and 1.41. This indicates that there were no extreme outliers because the critical value ($p < 0.05$) showing whether an extreme value is not due to chance alone is |3,49| for the present analyses. The degrees of freedom will, however, be affected. One degree of freedom is lost for each extra subject added to the first in the pooled data set. Hence, instead of df = 46 for a correlation analysis, we have df = 41. It should be noted that the results of these analyses do not permit prediction on the individual level, because the interindividual variation is removed.

To analyze the effects of, for example, time of night on the different variables, parametric two-way (2 nights × 4 times of night) or three-way (2 nights × 4 times of night × 3 types of tests) ANOVAs for repeated measurements using raw data were performed. All p-values from repeated measures ANOVAs are given after Greenhouse-Geisser epsilon correction (G-G correction) (13).

RESULTS

Performance on both tasks deteriorated significantly, and all three rating scales showed significantly increased sleepiness with time of night. There were no differences whatsoever between the 2 nights; there were also no significant interactions (Table 1). Therefore, data from the 2 nights are averaged in Figures 2 and 3.

The changes over time of night for performance and for KSS ratings are shown in Fig. 2. (Because VAS ratings were very similar to KSS ratings, they were not included in Fig. 2). KSS ratings of sleepiness increased markedly across time of night, irrespective of whether the ratings were given before the KAT, the vigilance test, or the reaction time test. However, sleepiness was always lowest before the KAT and highest before the reaction time test, with ratings before the vigilance test in between. That is, the sequence in which the tests were presented in each test block was reflected in the ratings. This effect across tests was significant for both KSS and VAS (Table 1).

The eight symptoms of the ATS scale were averaged to form a single score because an ANOVA (three-way repeated measurements: 4 times of night × 3 types of tests × 8 symptoms) showed no differences between the symptoms depending on time of night or type of test. Figure 3 shows the sharp decline in alertness across time of night for ratings given after the KAT, the vigilance test and the reaction time test.

Highest levels of ATS ratings appeared during the vigilance test and the lowest appeared during the KAT. This difference across tests was also significant (Table 1). Note that the order of ATS ratings was not the same as that for the KSS and VAS data.

The above data all show a clear decrease in performance and a clear increase in self-rated sleepiness as a function of time of night.

Relations between performance and self-ratings

Table 2 shows the results from correlation analyses (pooled data) where KSS and VAS ratings immediately before the vigilance test and the reaction time test, respectively, are used to predict actual performance on the respective tests. Clearly, both KSS and VAS ratings significantly predicted performance on both tasks. KSS explained more of the variation than did VAS. Both performance measures were highly related to the ATS ratings (Table 2). Degraded performance was associated with increase in symptoms of sleepiness.

Relations between rating scales

The relations between the three types of ratings are studied by relating the ATS rating after the KAT test

TABLE 3. *Relations between ATS and KSS/VAS ratings in close temporal connection (top) and relations between the KSS and VAS ratings given together before the two performance tests (bottom). Coefficients of correlations are based on standardized and pooled data from six subjects and eight measurements (n = 48). p-values with df = 41 (see text)*

	KSS(Vig)	VAS(Vig)	KSS(RT)	VAS(RT)
ATS(KAT)	0.73 ($p < 0.0001$)	0.70 ($p < 0.0001$)	—	—
ATS(Vig)	—	—	0.81 ($p < 0.0001$)	0.65 ($p < 0.0001$)
KSS(Vig)	—	0.86 ($p < 0.0001$)	—	—
KSS(RT)	—	—	—	0.71 ($p < 0.0001$)

Vig = vigilance task; RT = reaction time task.

to the ratings from KSS and VAS given together before the vigilance test, and by relating the ATS after the vigilance task to KSS and VAS given together before the reaction time task. The relations between the KSS and VAS ratings that were given together before the vigilance and reaction time tests, respectively, were also studied. Coefficients of correlation are shown in Table 3.

There were highly significant relations between the ATS ratings and the other two scales. The correlation coefficients ranged between 0.65 and 0.81. Moreover, scores on the KSS and VAS were highly positively correlated with each other before both performance tests.

DISCUSSION

There were clear effects of sleep deprivation on all variables. This effect was expected for the types of tasks used and from previous experience with KSS and VAS, but appeared also for the new ATS.

Ratings predicted performance to a high degree; 35–50% of the variance in performance scores could be explained by the prior KSS ratings, and this was slightly better than for the VAS ratings. In recent studies Johnson et al. (7,8) found that the SSS did not predict "objective" sleepiness measures (multiple sleep latency test and lapses in a tapping task). However, their design and methods for analyzing data might not have been optimal for the purpose of demonstrating relationships. The correlations were calculated across individuals on data aggregated over several measurements, a method that might adequately reveal individual differences but not the sensitivity of self-rating methods. Furthermore, there might have been a restriction of variance due to inadequate manipulation of the underlying sleepiness. Similarly, Herscovitch and Broughton (6) found no correspondence between SSS and performance. The temporal relations between ratings and performance were not clear, however, with the exception of ratings given immediately after tests.

There was an even closer covariation between the quality of performance and the ensuing rating of symptoms of sleepiness (ATS). The explained variances were around 50–60%. Glenville and Broughton (5) reported similar correlations for SSS with vigilance and reaction time performance. The higher correlations after tests might be explained by the fact that it may be easier to rate with precision after a specific task because the situation limits extraneous (uncontrolled) influences on alertness and, furthermore, because subjects may be influenced by their perceived quality of performance when rating their sleepiness. These explanations are supported by findings that show higher self-rated sleepiness during or immediately after a performance task (14–16). Although the present design does not allow the conclusion that ATS is a more sensitive "posttest" rating scale than KSS or VAS, one might certainly conclude that it is sensitive.

When we controlled for time of night, scores on the ATS indicated higher sleepiness after the vigilance task than after the reaction time task. This shows ATS to be sensitive to the dearousing effects of the task preceding the ratings. The vigilance task was more monotonous, demanded less action and was considerably longer than the reaction time task. In contrast to this, when time of night was controlled, the KSS ratings given before the tests retained the same order in the level of sleepiness as the order in which they occurred in each block.

The ATS ratings showed high correlations with the more traditional ratings, although they differed considerably in the cognitive processes required to perform the ratings task. There was a high consistency in how the subjects used the different rating scales. All three rating scales sensitively reflected the underlying sleepiness.

To conclude, there was a strong relation between ratings of sleepiness—including those obtained with the new ATS—and performance. The ATS ratings were affected by the length and type of task, which underscores the importance of knowing the details of an experimental session when interpreting subjective ratings of sleepiness.

Acknowledgement: This study was supported by the Swedish Work Environment Fund.

REFERENCES

1. Torsvall L, Åkerstedt T. Sleepiness on the job: continuously measured EEG changes in train drivers. *Electroencephalogr Clinical Neurophysiol* 1987;66:502–11.
2. Kecklund G, Åkerstedt T. Sleepiness in long distance truck driving: an ambulatory EEG study of night driving. *Ergonomics* 1993;36:1007–17.
3. Torsvall L, Åkerstedt T, Gillander K, Knutsson A. Sleep on the night shift: 24-hour EEG monitoring of spontaneous sleep/wake behaviour. *Psychophysiology* 1989;26:112–8.
4. Åkerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. *Internat Neurosci* 1990;52:29–37.
5. Glenville M, Broughton R. Reliability of the Stanford Sleepiness Scale compared to short duration performance tests and the Wilkinson auditory vigilance task. In: Passouant P, Oswald I, eds. *Pharmacology of the states of alertness.* Oxford: Pergamon, 1979:235–44.
6. Herscovitch J, Broughton R. Sensitivity of the Stanford sleepiness scale to the effects of cumulative partial sleep deprivation and recovery oversleeping. *Sleep* 1981;4:83–92.
7. 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; 13:121–35.
8. Johnson LC, Freeman CR, Spinweber CL, Gomez SA. Subjective and objective measures of sleepiness: the effect of benzo-

diazepine and caffeine on their relationship. *Psychophysiology* 1991;28:65–71.

9. Åkerstedt T, Gillberg M. The circadian variation of experimentally displaced sleep. *Sleep* 1981;4:159–69.
10. Gillberg M, Åkerstedt T. Possible measures of "sleepiness" for the evaluation of disturbed and displaced sleep. In: Reinberg A, Vieux N, Andlauer P, eds. *Night and shiftwork—biological and social aspects.* Oxford: Pergamon, 1981:155–60.
11. Lisper H-O, Kjellberg A. Effects of 24-hour sleep deprivation on rate of decrement in a 10-minute reaction time task. *J Exp Psychol* 1972;96:287–90.
12. Puri ML. *Non-parametric techniques in statistical inference.* Cambridge: Cambridge University Press, 1970.
13. Greenhouse SW, Geisser S. On methods in the analysis of profile data. *Psychometrica* 1959;24:377–93.
14. Angus RG, Heslegrave RJ. Effects of sleep loss on sustained cognitive performance during a command and control simulation. *Behav Res Methods, Instruments and Computers* 1985;17:55–67.
15. Dinges DF, Orne MT, Orne EC. Sleepiness during sleep deprivation: the effects of performance demands and circadian phase. *Sleep Res* 1984;13:189.
16. Babkoff H, Caspy T, Mikulincer M. Subjective sleepiness ratings: the effects of sleep deprivation, circadian rhythmicity and cognitive performance. *Sleep* 1991;14:534–9.