A Pictorial Sleepiness Scale Based on Cartoon Faces

Claudia C. Maldonado, BSc(Hons); Alison J. Bentley, MBBCh; Duncan Mitchell, PhD

Wits Dial·A·Bed Sleep Laboratory, Brain Function Research Unit, School of Physiology, University of the Witwatersrand, Johannesburg, South Africa

Study Objectives: To develop a sleepiness scale devoid of semantic or geometric elements.

Design: Subjects were asked to rank in order 7 cartoon faces representing degrees of sleepiness. We used Thurstone’s scaling procedure to transform these rankings into an interval scale, which allowed us to eliminate 2 of the faces. The remaining 5 faces were ranked again using other subjects. In a validation study, subjects rated their perceived level of sleepiness using our scale and other sleepiness scales. Employed shift-workers and school-going children used our scale to assess its practical applicability.

Settings: Research and diagnostic sleep laboratories, pre-primary to tertiary institutions, shift-working industry.

Participants: Ethnically diverse healthy and sleep-disordered adults (n = 490), and school-going children (n = 345).

Measurements and Results: Our faces scale correlated with the Karolinska Sleepiness Scale (P < .05), the Stanford Sleepiness Scale (P < .04), and a visual analog scale measuring sleepiness (P < .0001). Shiftworkers showed a time-on-task effect on the evening shift (P < .0001) and a peak in sleepiness at 4:00 and 5:00 PM (P < .0001) on the night shift. Eight to 10 year old children appeared sleepier than older children throughout a school day (P ≤ .02) and became sleepier as the day progressed (P < .0001). We confirmed that our scale measures sleepiness uncontaminated by pain, anger, or happiness.

Conclusions: We have devised a sleepiness scale suitable for people too young or insufficiently educated to employ more-conventional scales. We envisage the scale being used for diagnostic, therapeutic, and research purposes.

Key Words: subjective sleepiness, sleepiness scale, children, pictorial scale, faces scale

Abbreviations: MSLT, Multiple Sleep Latency Test; KSS, Karolinska Sleepiness Scale; SSS, Stanford Sleepiness Scale; VAS, visual analog scale; ESS, Epworth Sleepiness Scale; ANOVA, analysis of variance

Citation: Maldonado CC; Bentley AJ; Mitchell D. A pictorial sleepiness scale based on cartoon faces. SLEEP 2004;27(3):541-8.

INTRODUCTION

SLEEPINESS IS THE CRAVING OR DESIRE FOR SLEEP AND IS THE PHYSIOLOGICAL DRIVE THAT PROMOTES THE ONSET OF SLEEP1 AND REFLECTS THE BALANCE BETWEEN SLEEP DRIVE AND AROUSAL.2 The degree of sleepiness may be estimated objectively using the multiple sleep latency test (MSLT), which measures the time to sleep onset evident in the electroencephalogram.3 The MSLT has been used extensively to measure the extent of daytime sleepiness in patients with sleep disorders, particularly narcolepsy, as well as sleep propensity in research settings.4,5 The MSLT is sensitive to partial sleep deprivation and the cumulative effects of sleep restriction.6 A modification of the MSLT, the Maintenance of Wakefulness Test, is used to quantify the ability to remain awake and has been used to evaluate treatment efficacy in patients with excessive somnolence.7 Although these electroencephalographic techniques provide the most accurate measurements of sleepiness, they are inappropriate for large-scale epidemiologic studies and some fieldwork.

Subjective perceptions of sleepiness are reasonable indicators of physiological sleepiness and may even precede performance decrements.8 Generally, we are able to recognize how sleepy we are, though misjudgment of sleepiness may occur with prolonged sleep restriction.9 Subjective measures of instantaneous perceived sleepiness, such as those reflected in the Karolinska Sleepiness Scale (KSS)10 and the Stanford Sleepiness Scale (SSS),10 are quick and simple to complete, provided that the subjects have some level of literacy and comprehension of what is required. Visual analog scales (VAS),11-15 which are simple, cheap, and easy to execute, also have been used to assess sleepiness.8,9,16,17 However, indicating a response on a VAS requires some understanding of geometry and the ability to map an abstract phenomenon onto a VAS. The concept of a scale, even one as simple as the VAS, may be foreign to children at a prelogical cognitive stage of development. Also, preliminary studies in our laboratory have indicated that poorly educated and, especially, rural people fail to understand the concept of a VAS.

Conceptually, the measurement of perceived sleepiness is a task similar to the measurement of perceived pain. Pictorial scales, using cartoon faces, are well established as instruments for measuring perceived pain in subjects and patients for whom semantic or geometric scales are inappropriate and, particularly, for children.18 We therefore designed a pictorial scale, using cartoon faces, to measure perceived sleepiness and tested it using, as a subject base, a diverse South African adult population that included many subjects whose home language was not English and South African preschool and junior-school children. We also validated the scale against established scales and assessed its utility in practical applications.

MATERIALS AND METHODS

Subjects and Ethical Considerations

A cohort of 835 subjects participated (Table 1). Separate groups of subjects were used for each phase of development and testing. Subjects for the ranking procedure were recruited from patients attending a diagnostic sleep laboratory and their guests, healthy adults employed at our university, and healthy children from local English-medium preschool care centers and junior schools. Subjects for the validation were recruited from patients attending the same diagnostic laboratory for daytime studies and healthy university students. The application phase consisted of subjects recruited from 2 English-medium junior schools and adult shift workers employed in the mining industry.

Each component of this project was approved by the Committee for Research on Human Subjects of the University of the Witwatersrand. Informed verbal consent was obtained from all subjects of consenting...
age and from school authorities for the children. For underage subjects, written informed consent was obtained from adults legally responsible for them.

### Experimental Procedures

#### Ranking and Reduction of 7-Face Scale

Seven cartoon faces depicting degrees of sleepiness were drawn by a specialist in medical graphic images. Particular attention was paid to the eyes, and the faces were devised such that they were devoid of gender and ethnic features. The cartoons included 4 faces showing various degrees of eye closure, yawning, a hand rubbing an eye, and a person asleep (Figure 1). The cartoon of each face (approximately 55 mm x 60 mm), mounted on a magnet, was attached to a portable metal display board (640 mm x 640 mm). This construction allowed us to position the 7 faces randomly in 2 dimensions to avoid any implied rank or order.

Subjects (Table 1) were told the purpose of the study, using terms appropriate for their age, language, and level of cognition. Subjects were asked their age and with what ethnic group they identified themselves or what their home language was. The procedure took approximately 5 minutes per subject. They were assured that the task was not a test, quiz, or game and that there was no right or wrong answer. Respondents were asked to rank the faces in order, from the face they considered to depict the person most awake to the person most sleepy. Each subject's response was anonymous, confidential, and free from outside interference. Once the faces were ranked, each subject was asked to look at the assembly of faces in their selected order and was given a chance to amend their ranking.

Thurstone's scaling procedure (see below) was used to transform the individual rankings to a consolidated interval scale for each group of subjects. On the basis of that analysis, 2 faces were eliminated from the original set. The

### Table 1—Characteristics of subjects

<table>
<thead>
<tr>
<th>Phase</th>
<th>Task</th>
<th>Source of subjects</th>
<th>Number</th>
<th>Age (mean ± SD or range)</th>
<th>Male, (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranking</td>
<td>Reduction of 7-face scale</td>
<td>Diagnostic sleep laboratory</td>
<td>109</td>
<td>44 ± 12</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Guests of patients</td>
<td>36</td>
<td>40 ± 14</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University</td>
<td>16</td>
<td>34 ± 10</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Junior school</td>
<td>41</td>
<td>7 ± 1</td>
<td>100</td>
</tr>
<tr>
<td>Ranking</td>
<td>Reduction of 5-face scale</td>
<td>Diagnostic sleep laboratory</td>
<td>20</td>
<td>45 ± 14</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University</td>
<td>20</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preschool care center</td>
<td>20</td>
<td>5 ± 1</td>
<td>35</td>
</tr>
<tr>
<td>Validation</td>
<td>Comparison with KSS &amp; SSS</td>
<td>Diagnostic sleep laboratory</td>
<td>8</td>
<td>45 ± 11</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University</td>
<td>26</td>
<td>18 ± 41</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Comparison with VAS</td>
<td>University</td>
<td>234</td>
<td>22 ± 5</td>
<td>36</td>
</tr>
<tr>
<td>Application</td>
<td>Drivers</td>
<td>Mining industry</td>
<td>21</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Grade-3 school children</td>
<td>Junior school</td>
<td>137</td>
<td>8 ± 1</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Grade-7 school children</td>
<td>Junior school</td>
<td>147</td>
<td>13 ± 1</td>
<td>47</td>
</tr>
</tbody>
</table>

KSS, Karolinska Sleepiness Scale; SSS, Stanford Sleepiness Scale; VAS, visual analog scale

Subjects identified themselves as South African Caucasian (English-speaking and Afrikaans-speaking), South African Black (Tswana, Ndebele, Xhosa, Zulu, and Sotho), British, Portuguese, Chinese, Israeli, Indian, Polish, Italian, Turkish, Ethiopian, and Greek.

### Figure 1—Interval positions of the 7 original faces, calculated by the Thurstone scaling procedure, from the rank order selected by subjects in 4 groups, from the face that subjects considered to depict the person most awake to that which depicted the person most sleepy (total n = 202). A. Patients attending a diagnostic sleep laboratory. B. Patients’ guests who accompanied them to the sleep laboratory. C. Healthy schoolchildren aged 6 and 7 years. D. Ostensibly healthy adults employed by our university. The positions shown are relative to the face on the extreme left.
removing 5 faces (Figure 2A) were modified slightly according to comments made by some subjects.

**Ranking of 5-Face Scale**

The 5 faces were ranked again in order of perceived sleepiness, using different subjects, from the same sources. During this secondary ranking, in order for us to be sure that the younger children understood the concept of sleepiness, each child (aged 4 to 7 years) was asked to mimic somebody feeling sleepy or, if they were older, whether they knew what feeling sleepy meant. When presented with all 5 faces together, the youngest children seemed overwhelmed and often failed to respond to the instructions. We then presented them with 2 faces and asked them to place them on the board in order of sleepiness and added the faces 1 at a time until all the faces were ranked. The children then were permitted to revise the order if they so wished. Thurstone scaling again was used to transform the individual rankings into an interval scale for each subject group.

**Validation**

**Validation Against Other Sleepiness Scales**

Eight patients suffering from excessive daytime sleepiness who had been referred to a diagnostic sleep laboratory and 26 healthy university students paid honoraria for their participation in a comparison of our faces scale with the SSS and KSS. Patients and students were asked to rate their personal level of sleepiness or alertness using our faces scale (see Figure 2B), the SSS, and the KSS, with the scales presented in a random order on 4 occasions during a school day. The faces were presented in a linear horizontal array. Every hour, for the duration of a shift, each subject marked the face that best described how sleepy or alert he felt at that time. There were two 8-hour shift cycles over a 24-hour day, the evening shift beginning at 4:00 PM and the night shift beginning at midnight.

**Contamination with Other Affective States**

The same university students participated in a protocol designed to assess to what extent the faces scale might be contaminated by the intrusion of affective states other than sleepiness in perceptions of what the faces depicted. On the basis of a recent recommendation made in respect of faces scales employed to assess pain, we first modified our scale to eliminate the smile on the first 3 faces (neutral anchor, Figure 2C). Students were asked to rank our faces both with the smile anchor (Figure 2B) as well as the neutral anchor (Figure 2C) in order of intensity in various affective states, from most happy to most sad, from least pain to most pain, from least angry to most angry, and from as alert as ever to as sleepy as ever, with the possible contaminating states presented in random order. Due to the tedious nature of this task, not all the rankings were performed by every subject. Thurstone scaling procedure was applied to each affective state.

**Application**

**Subjects with Relatively Low Literacy Levels**

Three groups of subjects were used for this phase. Heavy-vehicle drivers engaged in shift work in the mining industry used our scale (Figure 2B) to gauge their level of sleepiness during their routine work. The faces were presented in a linear horizontal array. Every hour, for the duration of a shift, each subject marked the face that best described how sleepy or alert he felt at that time. There were two 8-hour shift cycles over a 24-hour day, the evening shift beginning at 4:00 PM and the night shift beginning at midnight.

**Children with Differing Ages**

Two groups of school-going children used our scale to rank their level of sleepiness on 4 occasions during a school day. The faces were presented in a random 2-dimensional array, at 8:00 AM, 9:30 AM, 11:30 AM and 12:30 PM. The children chose the face that best described how sleepy or alert they felt at that time. The assessments were administered by the teacher in charge of that class.

**Statistical Analyses**

Thurstone Case V scaling procedure converts rankings into a zero-based interval scale, in which the linear distance between adjacent items relates directly to the frequency of selection of the items in that order, rather than the reverse order. If items cluster on the scale, there is disagreement among subjects about their rank order, so they may be considered interchangeable. Using the interval value derived by the Thurstone procedure allows for the use of interval, rather than categorical methods of statistical analyses. VAS values were normalized, for statistical purposes, with the arcsin transformation. Pearson correlations were used to compare the Thurstone scores from the faces scale to the scores from the VAS, and repeated-measures Pearson correlations are

![Figure 2](image)
reported for the validation phase of the study. Most of the statistical tests employed the statistical package GraphPad InStat (Graphpad Software, Inc., San Diego, Calif). Student t tests were used to compare regression lines using the neutral and smile anchors, and $\chi^2$ statistics were used to compare the measurement of various affective states using a neutral or smile anchor on the faces scale. Probability values less than .05 were considered significant. Sample sizes are shown in each case.

**RESULTS**

**Ranking**

Characteristics of the subjects who participated are shown in Table 1. Subjects ranged in age from 4 to 73 years and came from a wide variety of backgrounds. For many, English was not their home language, but all could communicate adequately in English.

Figure 1 shows the Thurstone scores generated from the ranking of the faces by each of the 4 groups of subjects, when the original 7 faces were presented to 202 subjects, asked to rank them in order of increasing depicted sleepiness. Four of the faces, 3 of which represented stages of apparent wakefulness rather than sleep, were ranked in the same order by the 4 groups of subjects and were positioned in distinctly separate positions on the Thurstone scale by all groups except university staff. The remaining 3 faces, representing apparent states of drowsiness, were clustered on the Thurstone scale and were not ranked in the same order by the 4 groups of subjects. We therefore eliminated 2 of the 3 faces and, on the basis of spontaneous comments made by subjects, retained the face representing drowsiness by a hand rubbing the eye (Figure 2A). Figure 2B and 2C shows the same faces after the smile first was reduced and then was eliminated to reduce confusion between alertness and happiness (or other affective states).

Figure 3 shows the Thurstone scores generated by the responses of all 60 subjects (sleep-laboratory patients and healthy adults and children) ranking the remaining 5 faces in the improved format of Figure 2B. These 5 faces occupied distinct positions on the scale. Only 5% of the subjects ranked the faces in an order different from that of Figure 3.

**Validation**

Patients and university students assessed their personal level of sleepiness using our faces scale and both the SSS and KSS at 4 times spread over 1 day. However, no statistically significant circadian variation in sleepiness was evident in the SSS and was evident only in the student cohort for both our faces scale and the KSS (repeated measures analysis of variance [ANOVA] $n = 25$, $P < .05$, $F_{3,72} = 2.817$ for faces and $P < .05$ Fr (nonparametric) = 11.794 for KSS. We therefore compared the Thurstone scores of the 2 versions of our scale correlated significantly with each other ($r = 0.992$, $P = .0009$). We therefore asked subjects to arrange the faces as if they related to other states. The Thurstone scores resulting from attempts by university students to rank the faces in order of other affective states, in addition to sleepiness, are shown in Figure 6. The scale was nondiscriminatory except when used to assess sleepiness. Significantly more subjects were able to rank all 5 faces according to sleepiness than according to happiness, anger, or pain (Table 2).

**Application**

The sleepiness of heavy-vehicle drivers employed in the mining industry, during different shift cycles, is shown in Figure 7. The drivers simply indicated which face best represented their instantaneous sleepiness at every hour of the shift, and we then calculated sleepiness scores from the Thurstone interval value for those faces. Perceived sleepiness increased progressively during the evening shift toward midnight and peaked at 4:00 AM and 5:00 AM on the night shift. Maximum perceived sleepiness was not significantly different between the evening and night shift. The drivers were able to implement the procedure of selecting a face without disrupting their work schedule and delivered a full data set, our analysis of which delivers a plausible record of the sleepiness trend in the drivers. The data also illustrate the value of Thurstone scaling in converting categorical to interval data, amenable to further analysis.

The drivers constituted a group of subjects precluded from using semantic or geometric scales by virtue of their limited education. School children, some of whose ages precluded them from using more sophisticated scales, also were able to relate to our faces scale and to indicate their sleepiness during the school day by picking a face (Figure 8). On all 4 occasions during the school day, the younger children chose faces that indicated that they were sleepier than the older children ($t$ test, $P \leq .02$). The younger children reported progressively increasing sleepiness as the day progressed, whereas the older children were sleepier only at 12:30 PM. Even so, both groups of children reported relatively low levels of sleepiness and never reached the levels of sleepiness reported by
the vehicle drivers. Again, the faces scale delivered plausible assessments of sleepiness in a population too young to employ other scales.

**DISCUSSION**

Subjects differing widely in age and ethnicity were able to relate cartoon faces to perceptions of sleepiness. They could rank 5 faces consistently in order, from the face representing the greatest wakefulness to that representing sleep. The cognitive complexity required for this ranking task was low. There was no semantic element requiring sophisticated comprehension of English or any other language. Indeed, the subjects recruited to develop our scale spoke at least 16 different home languages. The geometric skills required were confined to an ability to rank in order, a skill considerably simpler than the mapping required when employing a VAS. Children in the age range of 4 to 7 years, with little or no reading or writing skills and limited vocabulary, were able to carry out the task successfully. Presenting the cartoon faces in a random 2-dimensional array and requiring the subjects then to arrange them in a single dimension eliminated possible observer-imported bias on the order subsequently chosen.

Our faces scale showed good agreement with established subjective sleepiness scales, such as the SSS and the KSS, more so in somnolent than healthy populations, and has the added advantage that no translation or explanation is required. Our scale also correlated well with positions...
on a VAS that covered nearly three quarters of the spectrum between as alert as ever and as sleepy as ever. Our scale, though, does not require the abstract task of mapping a perception onto a geometric scale as is required with the VAS.

Thurstone scaling was employed to convert individual rankings by a group of subjects into an interval scale representing both the group’s consolidated rank order and the degree of agreement about that order among the members of the group. That analytic approach initially allowed elimination of the 2 redundant faces from the assembly, reducing the assembly to 5 elements. Thurstone scaling has another advantage: the position of each face on the scale is identified by a numeric distance from the anchor point of the scale (which we chose to be the face depicting a state of greatest alertness).

Applying our faces scale in an occupational setting showed its potential industrial usefulness. The well-documented increased sleepiness observed on the night shift, particularly from 4:00 AM to 6:00 AM, is reflected in our scale, and the increased sleepiness associated with a time-on-task effect also is reflected in drivers’ responses on our scale in the evening shift. Children, aged 7 to 14 years, could relate to our cartoon faces describing various states of sleepiness, and analysis of the faces that they chose delivered plausible assessments of sleepiness, known to be low in preadolescents, particularly the maximum levels of sleepiness noted at 12:30 PM, and of how sleepiness progresses over the school day. Even though the younger children in our study indicated greater perceived levels of sleepiness than did the older children, this phenomenon may very well reverse as the children mature into adoles-

Table 2—Percentage of subjects who were able to rank all 5 faces according to sleepiness and other affective states, in the versions without a smile (Figure 2C)

<table>
<thead>
<tr>
<th>State</th>
<th>Sample Size</th>
<th>Able to rank (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleepiness</td>
<td>232</td>
<td>99</td>
</tr>
<tr>
<td>Happiness</td>
<td>157</td>
<td>48</td>
</tr>
<tr>
<td>Anger</td>
<td>111</td>
<td>34</td>
</tr>
<tr>
<td>Pain</td>
<td>118</td>
<td>52</td>
</tr>
</tbody>
</table>

Table 2—Percentage of subjects who were able to rank all 5 faces according to sleepiness and other affective states, in the versions without a smile (Figure 2C)

Significantly more subjects could rank the faces according to sleepiness than according to the other affective states \( \chi^2(3) = 195.6, P < .0001 \)

Figure 6—Thurstone scaling of the rankings of the five faces from the set with smile absent (Figure 2C) \((n = 232)\) in order of perception of sleepiness (D) and other affective states: A. Ranked from least pain to most pain. B. Ranked from least happy to most happy. C. Ranked from least angry to most angry. Faces tended to rank in the same order for all affective states, but their positions were distinct only for perceptions of sleepiness.
ence, and the rise in physiological sleepiness associated with pubertal development\textsuperscript{17,28} may be reflected in our faces scale. Using our faces scale in an occupational setting and in a school setting caused minimal disruptions to work and allowed an excellent response rate.

Attempts by educated subjects to rank our faces according to other affective states, namely pain, happiness, and anger, showed that it is a poor measure of these affective states. Almost all (99\%) of the subjects were able to rank the faces according to sleepiness, whereas significantly fewer subjects were able to rank our faces scale according to happiness, anger, and pain, irrespective of the anchor used. Thus our scale does not appear to be contaminated by the other affective states; the state reflected by choice of a face indeed is sleepiness.

We envisage the faces scale being used for diagnostic, therapeutic, and research purposes. Patients or subjects can be presented with the 5 cartoon faces in a random arrangement in a 2-dimensional or linear array and asked to point out the single face that best represents their current perceived state of sleepiness. An equivalent procedure, employing a different set of cartoon faces, has been used successfully in both therapeutic and research settings for the measurement of perceived pain in children, even as young as 3 years.\textsuperscript{18,19,29} VAS, with pictorial anchors at each pole, have been used in sleep research (M. A. Carskadon, PhD, personal communication)\textsuperscript{37} but still require subjects to master the geometric mapping. We have taken the functionality of our scale beyond that afforded by simple choice of a face by employing the Thurstone procedure to convert the ranking of the faces, in order of sleepiness, to an interval scale, thereby generating a number attached to each face, without assuming linearity. Other subject or patient groups could rank the faces and so establish their own Thurstone scores, or the scores generated by our subject groups could be employed. Thurstone scaling generates values that depend on sample size as well as degree of agreement, but the responses of different subject groups each with its own Thurstone scale still may be compared, provided the scales correlate with each other.

Our scale cannot have the sensitivity of more-sophisticated measures of sleepiness, and we do not expect it to replace these measures. Electroencephalographic measures of sleep onset (for example, the MSLT or nighttime sleep recordings) are likely to remain the most reliable measures of sleep propensity. Also, the MSLT continues to show decreased sleep latencies with accumulating sleep loss, whereas subjective perceptions of sleepiness (such as the SSS and VAS) plateau with accumulating sleep loss,\textsuperscript{3} and our scale may well do so, too. Although the MSLT is a highly reliable measure of daytime sleepiness in healthy normal subjects, provided there are more than 3 nap sessions,\textsuperscript{39} it and other electroencephalographic techniques\textsuperscript{9} are time consuming and expensive to perform. Also, patients with insomnia and those with mild obstructive sleep apnea have variable sleep latencies determined largely by the quality of sleep on the preceding night.\textsuperscript{30}

The KSS,\textsuperscript{9} a useful subjective assessment of sleepiness, correlates with the VAS\textsuperscript{3} and has been used successfully in research settings.\textsuperscript{9,25} The SSS is sensitive to sleep deprivation and correlates with performance measures\textsuperscript{8,10} and also is sensitive to circadian dips in performance.\textsuperscript{8} The ESS,\textsuperscript{30} which records recent sleep-related history, is reliable over time and internally consistent and distinguishes normal adults from patients suffering from excessive sleepiness.\textsuperscript{31} It cannot be used, however, to detect instantaneous variations in sleepiness. The ESS correlates well with nighttime polygraphic measures of sleep.\textsuperscript{32} Although some authors have shown ESS scores to be modestly to highly reliably correlated with MSLT scores,\textsuperscript{32,34} other authors have found no correlation between the ESS and MSLT scores in patients suffering from sleep-disordered breathing or periodic limb movement disorder.\textsuperscript{35} The ESS has been used extensively in clinical and research adult populations and distinguishes sleepy patients from healthy controls.\textsuperscript{30,31,33} The ESS has been used successfully in Italian and Portuguese subjects, (M Casagrande, C Violani, P Testa and G Curcio; A Pedrosu, CG Rose, P Kadre, et al, reported at the meeting of the Associated Professional Sleep Societies in San Francisco, Calif., June 1997) although translating and back- translating the scale, which is the minimum requirement for translating semantic scales, is time consuming and demanding.

The commonly used subjective sleepiness scales were devised for the more-literate adult, but children may experience difficulty with the abstract task of describing the exactness of subjective experiences using verbal cues.\textsuperscript{12} Children do better matching internal states with pictorial representations of subjective states. A simple, quick, and inexpensive measure of subjective sleepiness also has potential application for assessing degree of sleepiness in industry, with little disruption of work. A pictorial scale will therefore be particularly beneficial for use in illiterate or non-English-speaking workers and in children.

Our proposed pictorial scale is simple, quick to use, and requires only limited instruction. We predict the pictorial scale could be used to mea-

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{Sleepiness (mean \pm SD) of heavy-vehicle drivers (n = 9 evening shift, n = 12 night shift) at hourly intervals over two 8-hour shift cycles. The sleepiness scores were calculated by assigning the Thurstone interval value of each face in Figure 3 to the face chosen by the driver to best represent his sleepiness at the time, with low scores indicating greater wakefulness. The drivers on the evening shift showed significantly increasing sleepiness towards midnight (repeated measures analysis of variance, F\textsubscript{8},88 = 13.49, P < .0001). A peak in sleepiness occurred at 4:00 AM and 5:00 AM for the drivers on the night shift (repeated measures analysis of variance, F\textsubscript{8},88 = 5.61, P < .0001). Maximum sleepiness on this scale = 5.58.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure8.png}
\caption{Sleepiness (mean \pm SD) of Grade 3 (age 8.4 \pm 0.5 years, clear squares) and Grade 7 (age 12.5 \pm 0.7 years, solid circles) school children showing degree of sleepiness over a school day. Sleepiness scores were calculated from the faces that the children chose to represent instantaneous sleepiness, as in Figure 7, with low scores indicating greater wakefulness. The Grade 3 children reported sleepiness that increased progressively as the day progressed (repeated measures analysis of variance, F\textsubscript{3},544 = 13.76, P < .0001), whereas the older children were significantly more sleepy only at 12:30 PM (repeated measures analysis of variance, F\textsubscript{3},438 = 4.25, P < .0001).}
\end{figure}
sure subjective sleepiness in children as young as 4 years. This tool also could be used for epidemiologic studies or fieldwork and wherever verbal or cognitive limitations hamper communications. The sensitivity of the pictorial scale to sleep disruption and other interventions known to increase sleepiness, and its utility in tracking improvements in patients treated for sleep disorders causing excessive daytime sleepiness, are still to be explored. The ultimate usefulness of the scale will depend on the diversity of situations in which it can be applied successfully. We encourage other interested researchers and sleep medicine professionals to test the scale in their particular fields of interest; copies of our cartoon faces are available on request.

ACKNOWLEDGEMENTS

We thank Terry Borain of the Photo Illustration Unit of the Faculty of Health Sciences, University of the Witwatersrand, for the design of the faces; Willie De Bruin for constructing the test array; Yvonne Glazer for assisting with data collection in the patient populations; Leanne Katzenellenbogen for collecting some of the data of Figure 4; Alison Saunders, Brendon Billings, Praneshri Pillay, and Stacey Ramsey for collecting some of the data of Figure 4, 5, and 6; Schu Schutte for arranging the data collection for Figure 7; the teachers involved in the various phases of the study; and Helen Driver for comments on the manuscript. Our research was funded by the University of the Witwatersrand.

REFERENCES


SLEEP, Vol. 27, No. 3, 2004