Performance Criteria and Testing Protocols for Features in Sleep Tracking Consumer Technology Devices and Applications

ANSI/CTA/NSF-2052.3

April 2019

NOTICE

Consumer Technology Association (CTA)™ Standards, Bulletins and other technical publications are designed to serve the public interest through eliminating misunderstandings between manufacturers and purchasers, facilitating interchangeability and improvement of products, and assisting the reader in selecting and obtaining with minimum delay the proper product for the reader's particular need. Existence of such Standards, Bulletins and other technical publications shall not in any respect preclude any member or nonmember of the Consumer Technology Association from manufacturing or selling products not conforming to such Standards, Bulletins or other technical publications, nor shall the existence of such Standards, Bulletins and other technical publications preclude their voluntary use by those other than Consumer Technology Association members, whether the standard is to be used either domestically or internationally.

> *Standards and Publications are adopted by CTA in accordance with American National Standards Institute (ANSI) patent policy. By such action, neither CTA nor ANSI assumes any liability to any patent owner, nor does either organization assume any obligation whatever to parties adopting the Standard or Publication. CTA and ANSI take no position with respect to the validity of any Essential Patent Claim relating to this standard. Neither CTA nor ANSI is responsible for identifying patents for which a license may be required in order to comply with any CTA or ANS standard.*

This document does not purport to address all safety problems associated with its use or all applicable regulatory requirements. It is the responsibility of the user of this document to establish appropriate safety and health practices and to determine the applicability of regulatory limitations before its use.

(This document was produced by CTA's **R11 Health, Fitness & Wellness Committee**.)

Published by
©CONSUMER TECHNOLOGY ASSOCIATION 2019
Technology & Standards Department
www.cta.tech

FOREWORD

This standard was developed by the Consumer Technology Association's Health, Fitness and Wellness Technology Committee.

(This page intentionally left blank.)

CONTENTS

(This page intentionally left blank.)

Performance Criteria and Testing Protocols for Features in Sleep Tracking Consumer Technology Devices and Applications

Contents

1 Scope

This standard address performance criteria and testing protocols for features in sleep tracking consumer technology devices and applications.

2 References

2.1 Normative References

The following references contain provisions that, through reference in this text, constitute informative provisions of this standard. At the time of publication, the edition indicated was valid. All standards are subject to revision, and parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent edition of the standard indicated below.

2.1.1 Normative Reference List

- ANSI/CTA/NSF-2052.1, *Definitions and Characteristics for Wearable Sleep Monitors*
- ANSI/CTA/NSF-2052.2, *Methodology of Measurements for Features in Sleep Tracking Consumer Technology Devices and Applications*

2.2 Informative References

The following references contain provisions that, through reference in this text, constitute informative provisions of this standard. At the time of publication, the edition indicated was valid. All standards are subject to revision, and parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent edition of the standard indicated below.

2.2.1 Informative Reference List

- Blake H, Gerard RW, Kleitman N. Factors influencing brain potentials during sleep. J Neurophysiol 1939; 2: 48-60. [1]
- Czeisler CA, Buxton OM. Human circadian timing system and sleep-wake regulation, Sixth Edition. In: Kryger MH, Roth T, Dement WC (Eds) Principles and Practice of Sleep Medicine. Elsevier: Philadelphia; 2017, 362-376.[2]
- Benloucif S, Burgess HJ, Klerman EB, Lewy AJ, Middleton B, Murphy PJ, Parry BL, Revell VL. Measuring melatonin in humans. J Clin Sleep Med 2008 ; 4(1): 66-69.[3]
- American Academy of Sleep Medicine. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications. www.aasm.org. American Academy of Sleep Medicine: Darien, IL; 2014. [4]
- Mang GM, Nicod J, Emmenegger Y, Donohue KD, O'Hara BF, Franken P. Evaluation of a piezoelectric system as an alternative to electroencephalogram/ electromyogram recordings in mouse sleep studies. Sleep 2014; 37(8): 1383-92. [5]
- Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages in human subjects. NIH Publication No. 204. Washington: U.S. Government Printing Office; 1968. [6]
- Iber C, Ancoli-Israel S, Chesson A, Quan SF. For the American academy of sleep medicine: The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Westchester, Ill: American Academy of Sleep Medicine; 2007. [7]
- Measurement and Observer Effect. Citation found using Google on Aug 28, 2018. https://www.intropsych.com/ch01_psychology_and_science/measurement_and_observer_effects.html [8]

2.3 Compliance Notation

As used in this document “shall” and “must” denote mandatory provisions of the standard. “Should” denotes a provision that is recommended but not mandatory. “May” denotes a feature whose presence

does not preclude compliance, and implementation of which is optional. "Optional" denotes items that may or may not be present in a compliant device.

2.4 Definitions

Test Device (TD) is the device being validated in this standard, for example, a wearable sleep monitor.

Gold Standard (GS) is a previously validated device or technology or methodology against which the TD is compared, for example, video in the context of Time in Bed (TIB), or clinical polysomnography (PSG) in the context of sleep/wake classification.

3 Preface

The following serves to facilitate understanding of this document focused on performance criteria for evaluating wearable and in-bedroom devices designed to monitor sleep. The scope is intentionally limited to "elemental" measures as defined in the previous document entitled Definitions and Characteristics for Wearable Sleep Monitors ANSI/CTA/NSF-2052.1. Other measures are derived from these measures and will reflect accuracy or inaccuracy of the elemental measures; therefore, attention centers on precision and quantitative correctness of elemental parameters. The relevant measures are: Time Attempting to Sleep (TATS) Start Time, TATS End Time, Time in Bed (TIB) Start Time, TIB End Time, Awake, Asleep, Awakening from Sleep, Brief Awakening, Initial Sleep Onset Time, Final Awakening Time, Brief Moment of Sleep (Dozing), REM Sleep, N1 Sleep, N2 Sleep, N3 Sleep, CNS Arousal, Dream Sleep, Core Sleep, Restless Sleep, Sound Sleep, Circadian Amplitude, Circadian Period Length (tau), and Circadian Phase.

Table 1 summarizes the measures relevant to this standard, their type (see below) and methods used to evaluate wearable and in-bedroom sleep monitoring devices.

The measures fall into three main groupings with respect to the type of parameter they represent. This is important because different types of parameters require different evaluation metrics. The approach to determining accuracy for a specific **event** (e.g., the time at which something occurred) differs from verifying that an ongoing **process** within a specific time domain is occurring concomitantly. Furthermore, methods will also differ when attempting to compare **patterns** recorded by two different measuring instruments over even longer time periods. The following generalizations apply in this regard:
Events are evaluated in terms of their variance from the measurement considered as a "standard". In some cases, the standard may be measured **directly** (e.g., observing a video in showing a person getting in bed). However, sometimes measures should only be evaluated indirectly by **correlative methods** (e.g., using concurrent PSG to determine if REM sleep is present).

Statistical methods differ for assessing a device's accuracy depending on the nature of the measure under scrutiny; that is, whether it is an (a) event, (b) process, or (c) pattern.

3.1 Evaluating Events

If an event is unique during a recording session (e.g., Initial Sleep Onset Time) then by averaging across trials (nights) and subjects (participants) one may get a mean of the absolute difference between the standard used and the device being evaluated. To avoid having outliers skew the overall distribution of scores and exaggerate differences that may occur in a small minority of cases, a "tolerance" may be set appropriate to the parameter and a binary determination whether the event was successfully determined within the stated limits. The calculation will be for the percentage of nights (or subjects) that the event was correctly identified. If an event occurs multiple times (potentially) during a recording session (e.g., Awakening from Sleep), then an intra-subject and inter-subject mean must be calculated separately.

3.2 Evaluating Processes

Processes related to sleep are considered within a time domain (e.g., 30 seconds) rather than as a momentary event. Although generally referred to as a "sleep state", the underlying activity is in constant flux and is categorized as one "state" or another depending on the occurrence of an electrophysiological event (e.g., a sleep spindle) or based on the activity present during the majority of the specified time. Thus, this categorization represents a generalization about the underlying process. Correlative methods

may be used to determine the extent of agreement between some designated standard (e.g., PSG) and the categorization made by the device being tested. While not as convincing as using a direct method, correlative methods may be the only ones available (or sometimes much more expedient).

3.3 Evaluating Patterns

Evaluating patterns often presents difficulty and may require different analytic techniques for different measures. Fortunately, in this case the measures are few (Circadian Amplitude, Circadian Tau, and Circadian Phase) and may be compared with an extant standard (core temperature). Circadian amplitude analysis is only applicable for devices measuring body temperature and may be compared directly for peak-to-trough agreement within each 24-hour cycle and then averaged over several days. By contrast, Tau may be determined for the sleep-wake cycle and then compared to the standard. Circadian phase may be determined by autocorrelation to determine the sleep-wake cycles offset from, for example, the internal biological clock as determined by melatonin release or core body temperature pattern. The autocorrelation should compare the standard to the device being evaluated. The two curves should be compared for the entirety of their recording by mathematically temporally shifting one in small time increments for a 24-hour period (plus and minus 12 hours). The wrap-around technique using the 24-hour period should be used as the recording is shifted. The point of maximum correlation provides an estimate of the phase offset between the curves.

3.4 Methods for Performance Evaluation: Toolbox

3.4.1 Direct Methods

The following are considered Direct Methods for evaluating device performance: 1) direct observation, 2) self-report, 3) direct observation with self-report, 4) Kleitman's direct observation paradigm [1], 5) self-report upon awakening, and 6) core body temperature (for circadian measures). Each of these methods is described in detail in the following paragraphs.

3.4.1.1 Direct Observation (DO)

Video recordings shall be made of the individual during testing to serve as source data. Multiple raters (two or more) shall review the video recordings independently to determine when events or processes have occurred. Disagreements between raters should be resolved before comparisons are made to evaluate the performance of the test device. Parameters for which DO may be used to evaluate sleep-tracker performance include: 1) TIB Start Time, 2) TIB End Time, 3) Awake During Active Wakefulness, 4) Final Awakening Time.

3.4.1.2 Self-Report (SR)

SR is when an individual tells the tester about their status. The individual should do so orally or using some indicator (e.g., a microswitch or touch sensitive device taped to their hand) to signal their status. If a non-verbal indicator is used, it should be unique enough to distinguish accidental signals (e.g., three button presses to signal having awakened from sleep). Video recordings should be made of the individual during testing to serve as source data. When self-report signaling is used, signals should be time-locked to video recordings. Parameters for which SR should be used to evaluation sleep-tracker performance include: 1) TATS Start Time, 2) TATS End Time, and 3) Final Awakening Time.

3.4.1.3 DO with SR

In some circumstances DO is not sufficient, by itself, to determine an individual's sleep-wake status. In such cases, SR is needed for confirmation. This is the case for Awakenings from Sleep that are not accompanied by sustained movement or commencement of purposeful behaviors. In such cases, the subject signals orally or instrumentally that they are indeed awake. The shorter the duration of the wakefulness, the more SR is needed to verify an awakening has occurred.

3.4.1.4 Self-Report Upon Awakening (SRUA) Paradigm

The SRUA Paradigm is a useful tool to verify occurrence of three specific processes. The first is an individual actually being asleep when inactive. The SRUA paradigm is an intervention approach in which the sleeper is queried to determine through SR whether he or she was actually asleep after being intentionally awakened (from presumed sleep) and asked, "were you asleep just now or awake?" The interrogator must not know the status according to the sleep-tracker when the SRUA paradigm is applied

to avoid unintentional or subtle biasing during questioning. The second parameter appropriate for testing with SRUA paradigm is Dream Sleep. The interrogator asks, “were you dreaming just now?” Again, the questioning is performed blind to knowledge of the test participant’s sleep status. Finally, the third process an SRUA paradigm should be used to evaluate is Core Sleep. The interrogator askes both the “were you asleep just now?” and “were you dreaming just now?” If the subject affirms question 1 with a “yes” and answers “no” to the second question, this provides evidence that Core Sleep was present. If the subject is “unsure” then it should be scored as sleep.

3.4.1.5 Kleitman’s DO paradigm

Nathanial Kleitman [1] conducted studies to determine the EEG correlates of sleep onset. Because of the known relationship between muscle relaxation and sleep onset he had subjects hold a light spool between two fingers as they were falling asleep. Dropping the spool was taken as the marker for sleep onset and concurrent changes in EEG pattern were described. This approach has been modernized and devices using micro-switches (or similar devices made with pressure sensitive materials) appear may be devised or purchased. EEG alpha disappearance ranged from 0.5 to 25 seconds before the spool was dropped. Polysomnographic criteria for sleep onset are largely defined by the disappearance of EEG alpha activity; consequently, for equivalent precision the same time range should be used to judge performance of a wearable or in-bedroom device.

3.4.1.6 Core Temperature (CT)

Core body temperature is the original, classic measure used to define circadian rhythms. It involves recording core temperature continuously with samples taken frequently enough to track the cyclic pattern. The original studies [2] used continuous analog measures but later research finds that sampling at much slower rates is adequate (e.g., 10-minute intervals). The original approach for measuring circadian rhythm relied on a rectal probe. This approach is still considered the “gold standard”; however, alternative technologies evolved to provide less invasive approaches. These include (1) the “Drager” double sensor that derives CT from heat flux gradients, (2) a swallowable temperature transmitting probe (gastrointestinal “radio pill”), (3) ear (tympanic membrane) temperature, (4) other infrared devices, and (5) subcutaneous embedded temperature transmitting probes.

3.4.1.7 Dim-Light Melatonin Onset (DLMO)

DLMO may be used to assess the timing of the internal biological circadian clock. It currently is considered the most reliable measure of central circadian timing in humans [3]. The onset of melatonin secretion when measured in dim light conditions provides the onset phase timing of the circadian sleep-wake rhythm. Melatonin may be determined in blood plasma, urine and saliva. Saliva sampling is a practical and reliable method to assess circadian phase in field-based, laboratory, and clinical settings. It is recommended to start collecting saliva in 30- or 60-minutes intervals 6 hours before habitual sleep onset (preferred) or at least 3 hours before habitual bed time. The saliva samples should be taken under dim light conditions (<30 lux). Wearing blue blocking glasses during the full period 3 to 6 hours period might aid to the required limited light input to the eye in non-controlled settings. The DMLO is defined as the time at which salivary melatonin exceed and remains above a threshold of 3 pg/ml or the time the concentration is 2 standard deviations above the mean of the first 3 baseline samples (only applicable when using the 6 hours sampling). Saliva can be collected by either passive drooling in a vial or using absorbent oral swabs. Most melatonin assays require a minimum of 1ml of saliva per sample for duplicate analyses. Samples should be stored in a freezer as soon as possible after collection and should be stored at -20 °C until assayed [using enzyme-linked immunosorbent assay (ELISA) or radioimmunoassay (RIA)]. Rinsing of the mouth with water before saliva sampling is preferred and certain food restriction may apply to avoid contamination of the saliva samples.

Determining melatonin levels by either blood sampling or urine sampling require different protocols and different threshold values apply. Since blood sampling is an invasive technique, it might be applied in a clinical setting. Urine sampling has a lower temporal resolution and is considered less accurate for DLMO determination.

3.4.2 Correlative Methods

The following are considered Correlative Methods for evaluating device performance: 1) bed sensor, 2) PSG, 3) rapid eye movement sensor, 4) electroencephalograph (EEG) sensor, and 5) validated research actigraph (ACT). Each of these methods is described in the following paragraphs.

3.4.2.1 Bed Sensor

A bed sensor may be used to determine TIB Start Time and TIB End Time. The bed sensor needs to have contacts or sensors reacting to body pressure. The sensor should be below the hip or chest zone of the test person, not disturbing the comfort feeling around that section. The contact should be open when the person is not in bed, and closed when the person is in bed, or vice versa. The contact should react to pressure in a range of 1 to 5 kPa for a period of more than 5 seconds to record a person is in bed. Other technologies providing equivalent sensitivity may alternatively be used (e.g., piezo sensor). A system like this will also record the person getting up during the night, but in this context only the first action (test person in bed) and the last action (test person gets up) are decisive for TIB. The time system of the bed sensor must be synchronized to the time system of other devices used in this test.

Caveats:

1) A bed sensor is not capable of recording TATS Start Time or TATS End Time, because the person might be in bed not attempting to sleep (e.g., watching television, checking mails, using social media).
2) Bed sensors detecting the presence of a test person with other methods than pressure (e.g., detecting heart rate, movements) are not recommended as a validation test method in this context as there may be unknown tolerances detecting presence reliably.

3.4.2.2 Polysomnography (PSG)

Current PSG technique involves continuously recording of frontal (F), central (C), and occipital (O) electroencephalographic (EEG) activity throughout an entire major sleep period (usually overnight). Monopolar derivations from F4, C4, and O2, referenced to contralateral mastoid (M), serve as primary data. Homologous left-sided EEGs serve as backup in case primary signals become eroded or compromised during the many hours of recording. The American Academy of Sleep Medicine (AASM) guidelines [4] also permit use of an alternate recording montage that substitutes midline bipolar recordings from frontal and occipital derivations. Electrooculographic (EOG) recordings derive from electrodes placed near the eyes' right and left outer canthi, each referenced to a neutral site (usually a mastoid) and recorded on separate channels. One eye electrode should be placed 1 cm above and the other 1 cm below the outer canthus. Thus, lateral eye movements produce robust out-of-phase EOG activity as the eye's positive corneal potential moves toward one electrode and away from the other. This characteristic out-of-phase signature allows easy differentiation of eye movements from frontal EEG activity (presenting as in-phase activity at the same electrodes). Some appreciation of vertical eye movements is afforded by placing one electrode slightly above and the other slightly below each eye's horizontal plane. For clinicians wishing to better visualize vertical eye movements, an optional recording montage (with right and left eye outer canthi electrodes both placed 1 cm below the horizontal plane and referenced to the middle of the forehead) is permitted. Submentalis electromyographic (EMG) activity derives from an electrode pair placed 1 cm above (on the horizontal midline) and the other placed 2 cm below the mandible's inferior edge (2 cm to the right of midline). A backup electrode is placed 2 cm to the left of midline.

The recordings made in this manner are scored according to standardized rules provided in the AASM manual. Essentially, each 30-second epoch is classified as awake, stage N1, N2, N3 or R according to EEG, EOG, and EMG activity. Sleep scoring of the PSG recording should be executed by a certified RST or RPSGT, or a trained somnologist, sleep clinician or sleep researcher. Ideally all PSGs should be visually scored by two scorers independently, resulting in a single consensus scoring, based upon discussion and agreement of the 2 scorers on the differences in the sleep annotation made by the 2 scorers. A viable alternative it to visually score the PSG recording by a single scorer, followed by an epoch-by-epoch visual review of that first scoring by another scorer, that serves as reviewer. Based upon discussion and agreement to the observed differences in the sleep annotation between scorer and reviewer, a single consensus scoring will be reached. Only in case of limited availability to multiple trained

sleep scorers, the tester may opt for a visual scoring by a single trained sleep scorer or by automated sleep scoring followed by an epoch-by-epoch visual review of that automated scoring by a trained scorer.

3.4.2.3 Eye Movement (EM) Sensor

In sleep research and clinical laboratories, a small electrode is placed near (1 cm) the outer canthus of each eye. Each eye electrode is referenced to a neutral recording site (e.g., attached to the earlobe or pasted onto the mastoid behind one ear) and recorded on a separate channel. Because the eye's cornea has a high positive electrical potential, a lateral change in the direction of gaze will show increasing positivity in one channel and negativity in the other. To capture vertical eye movements, the electrodes are attached with one slightly above (1 cm) and one slightly below (1 cm) the two eyes' horizontal plane. An alternative method detecting eye movements involves using a piezo-electric sensor on the eyelid or embedded in an eye mask (blindfold). Piezo-electric motion sensors have performed well under controlled situations and therefore should be used to evaluate performance of more novel approaches [5].

3.4.2.4 Electroencephalograph (EEG) Sensor

EEG sensors are used in sleep and neurology clinical practice. For sleep studies, electrodes should be placed over the central and occipital lobes at scalp locations and referenced to an electrically neutral site. (the earlobe or on the mastoid behind the ear). The potential differences are amplified and filtered (low pass at 35 Hz and high pass at 0.3 Hz). EEG definitions for sleep are described in the Rechtschaffen and Kales (1968) [6] and the American Academy of Sleep Medicine standardized manual (2007) [7].

3.4.2.5 Validated Research Actigraph (ACT)

Sleep-wake patterns assessed by actigraphy, are largely influenced by the habits, the environment, and social activities. For this reason, when making comparisons of sleep-wake assessments using ACT, it is important that a device that is at least validated for sleep assessment. The ACT device must be FDA cleared to market and show overall agreement (sleep and wake combined) with PSG of at least 85% and specificity (the ability to correctly identify wakefulness) greater than 50% for a population of at least 20 subjects in at least one peer reviewed publication.

Parameters for which ACT may be used to evaluate sleep-tracker performance include the pattern types; however, to assess actual circadian parameters core body temperature and/or DILMO must be used. ACT devices can determine habitual sleep-wake cycles and determine their regularity or irregularity.

3.5 Sample and Testing Conditions

Testing shall include at least 32 participants. Larger samples are encouraged. Participants shall be equally represented by gender (50% male and 50% female). Eight, or more, participants' age should fall in the following groups: Young Adults (18-25 years), Adults (26-64), Older Adults (65 years or older)[1].

All participants should be healthy and normal. The initial test sample should not include untreated individuals diagnosed with sleep disorders, medical disorders, neurological disorders, or psychiatric conditions.

Each participant should have a total bed time not less than 420 minutes per night and each participant should have one or more sleep nights. Multiple nights are preferred to reduce error variance.

Sleep-tracker device performance should be evaluated in a sleep laboratory or a specially equipped room. Sleep rooms should be private, darkened, and sound attenuated. The temperature and humidity range during testing should be controlled within reasonably comfortable limits by the tester. The room

[1] Age groups developed in the NSF publication about sleep time duration recommendations are used here. These age groups were established by a panel of experts using the RAND Appropriateness Method (National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Hirshkowitz et al. Sleep Health 1 (2015) 40–43)

should minimally be equipped with total darkness video recording equipment and a two-way intercom system.

3.6 Notes and Commentary for Each Measure's Evaluation

3.6.1 TATS Start Time and TATS End Time

The actual time that an individual intends or begins to attempt to sleep and ends his or her intention to sleep is only knowable by truthful self-report. Self-report should take the form of a verbal statement or by the individual instrumentally signaling his or her intention (e.g., by pushing a button on the device). In a clinical sleep laboratory, lights-out and lights-on are taken as TATS Start Time and TATS End Time. However, this only works in the controlled situation and turning on or off the lights in one's sleep environment is suggestive but not certain in ambulatory humans wearing sleep-trackers. In fact, TATS start, and end times may occur outside the sleep environment altogether (e.g., while on public transportation).

3.6.2 TIB Start Time and TIB End Time

Once the bed sensor reacts (person in bed) this time needs to be recorded. In case the person gets up within a time frame of 15 minutes after first contact that recording is ignored and a new TIB Start Time is determined once the person gets back to the bed. In case the test person gets up after 15 minutes of first contact, the original recording prevails.

3.6.3 Awake, Asleep

The direct method of measurement is DO with SR. The correlative method is PSG.

When a person is active, speaking coherently or engaging in purposeful behavior are direct indicators of wakefulness. Self-report from an individual that he or she is awake may serve as a direct measure of wakefulness. A sleep tracker that is able to provide real-time monitoring would thus be easy to test by asking an individual at various times during the sleep period when wakefulness occurred spontaneously whether he or she was awake or asleep. In the absence of real-time monitoring capability, one may use direct observation via video link to time such a sleep-wake question during episodes of spontaneously occurring episodes of wakefulness embedded in a sleep period. One may also provide the test subject with a microswitch/pressure/touch sensitive pushbutton and instruct them to press the switch if and when they awaken during the night. Time-synched sleep tracker data may then be parsed for sensitivity and specificity, for both long and brief awakenings.

Concurrent recordings of a sleep-tracker and an alternative measure could be compared in a continuing time domain analysis to verify the tracker's ability to determine wakefulness's presence. One could also perform intervention trials of "awakening" an individual at various times during the night and asking them if they were awake or asleep; however, the distribution of sleep across the night is not equiprobable, i.e., normal individuals sleep 85% of their bed time (or more) on any given night. One may circumvent this event distribution difficulty by testing during a series of daytime nap opportunities to assure that the intervention is not always occurring in a background of sleep. This helps avoid the "I was awake" response being disproportionately over represented. Ten interventions per hour for an overnight study and 5 per hour on an hour-long nap should provide adequate data for analysis.

The above approaches to determining wakefulness or sleep use direct measures and are therefore advantageous. However, two issues should be considered. (1) Theoretically, it can be argued that one is disturbing sleep by interacting with the test individual. A widely appreciated principle [8] in science is the "observer effect" theory which states that simply observing a phenomenon necessarily changes that phenomenon. This may be the result of instruments or disruptions caused by the measurement technique. Thus, the phenomenon being measured is changed, may no longer be "normal", and may manifest in different ways. (2) Practically, in some circumstances it may be difficult, procedurally intensive, and impractical to evaluate performance in this interactive manner. Therefore, one may wish to relay on a correlative approach (as described below).

SR of being awake correlates well with EEG measures of sleep in normal healthy individuals who have distinct alpha brainwave activity when relaxed with their eyes closed. Also, many autonomic measures

show alterations at sleep onset and during sleep. Such physiological parameters may be useful to detect the presence or absence of sleep. These include respiration, heart rate, and/or blood pressure. Some devices (e.g., peripheral arterial tonography) show high sensitivity for detecting sleep. Actigraphy and PSG are used in sleep research and sleep medicine for estimating and determining sleep, respectively. Non-actigraphic measures of movement detected using in-bed or bedside monitors may be tested in an analogous fashion.

Percentage of agreement between a sleep-tracker's data and the self-reported, physiological, or actigraphic/movement-based data should be reported.

3.6.4 Awakening from Sleep

The direct method for determining awakenings from sleep are made by observation or SR (either orally or instrumentally). A correlative approach would use polysomnographic indicators of wakefulness (i.e., more than 15 seconds of EEG alpha activity arising from a background of ongoing N1, N2, N3, or REM sleep).

3.6.5 Brief Awakening

Brief awakenings should be determined in the same manner as regular awakenings from sleep; however, for the correlative approach only 3-15 seconds of EEG alpha activity are required. PSG measures may be preferable for this measure because very short central nervous system arousals are often insufficient to recruit volitional behaviors. Nonetheless, we know that an individual having multiple brief awakenings during sleep will feel less refreshed even if he or she does not remember awakening many times.

3.6.6 Initial Sleep Onset Time

Using Kleitman's DO paradigm one may determine the percentage of time the "spool" dropped or the indicator switch was released, and the tracker also indicated the first sleep onset within a plus or minus 25 second surround. Similarly, the percentage of time EEG alpha activity diminished adequately to classify a 30-second epoch as N1, N2, N3, or REM occurred within 25 seconds of when the sleep-tracker demarcated initial sleep onset should be reported.

3.6.7 Final Awakening Time

As with the evaluation approach for awakening from sleep, the direct method would use observation or self-report (either orally or instrumentally). A correlative approach would use polysomnographic indicators of wakefulness (i.e., more than 15 seconds of EEG alpha activity arising from a background of ongoing N1, N2, N3, or REM sleep). The difference here involves determining the FINAL awakening so that other metrics are able to be calculated (e.g., latency to arising which is the time from final awakening to TATS end time).

3.6.8 Brief Moment of Sleep (Dozing)

A performance criterion is not provided in this document but will be considered in future versions of this standard.

3.6.9 N1, N2, N3, and REM Sleep

There are no direct measures for sleep-tracker performance evaluation. These terms are defined according to sleep medicine and sleep research criteria. Consequently, polysomnographic correlative measures should be used to assess performance (see illustrative sample for guidance).

3.6.10 Dream Sleep

Dream sleep differs from REM sleep in that REM sleep is defined according to sleep medicine and sleep research criteria. Nonetheless, the original interest in REM sleep derived from the finding that this particular sleep process seemed to be the underlying biological substrate associated with dreaming. Subsequent research indicated that the eye movements during REM sleep occur in conjunction with the direction of gaze in the dream. Researchers have used PSG and other technologies to detect when dreaming is occurring, so a sleeper may be awakened and interrogated about their dream content. This approach, SR upon awakening, represents the only direct method to verify that dreaming is occurring. However, it does create an "observer effect" and depending upon the number of awakenings may disturb the overall sleep process.

Correlative measures include using EEG sensors, eye movement sensors, or full PSG. If PSG is used as the comparator, dream sleep should be evaluated against REM sleep. If a correlative approach is used, each 30-second epoch should be classified as either dream sleep or not dream sleep. The sleep tracker's performance should be evaluated based for sensitivity and specificity. Overall agreement rates should also be calculated.

3.6.11 Core Sleep

Core sleep is all sleep except for dream sleep. Direct observation may be used to rule out the presence of active wakefulness and SRUA may be used as a direct measure by asking the sleeper "were you awake or asleep". See dream sleep above for caveats about "observer effect". Correlative methods may use EEG sensors, eye movement sensors, or full PSG. If a correlative approach is used, each 30-second epoch should be classified as either core sleep or not core sleep. The sleep tracker's performance should be evaluated for sensitivity and specificity. Overall agreement rates should also be calculated.

3.6.12 Restless Sleep

Restless sleep refers to sleep associated with significant movement activity. Theoretically, direct observation could be used to detect movement activity; however, the results would likely be unreliable and impractical. Correlative approaches could use PSG or a validated research actigraph. Each 30-second epoch should be classified as either restless sleep or not restless sleep. The sleep tracker's performance should be evaluated for sensitivity and specificity. Overall agreement rates should also be calculated.

3.6.13 Sound Sleep

No direct measures are available to index sound sleep. An EEG sensor or PSG may be used to detect high amplitude, low frequency EEG activity. If PSG is used, the sleep tracker's performance should be evaluated against N3 sleep. If EEG is used without PSG, the sleep tracker's performance should be evaluated against EEG greater than 75 microvolts with a frequency less than 2 cps. Slow waves are a subset of EEG delta bandwidth activity occurring during sleep. Slow waves are slower and have a greater amplitude. Slow waves are also the defining characteristic of N3 sleep determined by PSG. Each 30-second epoch should be classified as either sound sleep or not sound sleep. The sleep tracker's performance should be evaluated for sensitivity and specificity. Overall agreement rates should also be calculated.

3.6.14 Circadian Amplitude

Circadian amplitude metrics depend upon measures used. In general, the amplitude is a peak-to-trough difference (apogee to nadir) for the measure used. The most commonly used measure is core body temperature. Core temperature is considered a direct measure because it was used as a primary approach for defining the circadian rhythm in research. To evaluate performance of a wearable device, the mean of the absolute differences between the standard used and the sleep-tracker should be reported along with a measure of its variability (e.g., coefficient of variation, standard error, or standard deviation).

3.6.15 Circadian Period Length (tau)

Tau provides the duration (in time) of one complete circadian cycle. Core body temperature is considered a direct measure inasmuch as it was a principle measure used to define human circadian rhythms. Examining several successive days of information can be helpful to accurately determine tau. In the case of continuous measures (e.g., core temperature) several curve fitting methods have been used to improve precision. If curve fitting techniques are used, the procedure should be specified. To evaluate performance, the mean of the absolute differences between the measure of tau by standard used and the sleep-tracker should be reported along with its coefficient of variation. Alternatively, standard error or standard deviation may be reported rather than coefficient of variation.

3.6.16 Circadian Phase (phi)

DILMO and/or core body temperature should be used as a direct measure for comparing a sleep tracker's indication of circadian phase. Ultimately, the calculation of phase depends upon the internal circadian

rhythm's relationship to some environmental measure (e.g., sundown or sunrise). In this sense, it is a derived pattern measure not derived from other circadian measures but rather by comparison to an external measure. The phase relationship should be expressed as time before or after the environmental event. For example, if core temperature drops to nadir preceded bed time by 100 minutes, it would indicate an advanced sleep phase. The phase relationship could be express more scientifically in geometrically in terms of the degrees of offset in relationship to the overall circadian rhythm. None the less, the key to performance accuracy is the agreement between the overall timing of the circadian rhythm according to the standard compared to that of the sleep tracker. To evaluate performance, the mean of the absolute differences between the circadian rhythm midpoint determined by the standard used compared to the sleep-tracker should be reported along with a measure of its variability (e.g., coefficient of variation, standard error, or standard deviation).

3.7 Illustrative Examples for Events, Processes, and Patterns

3.7.1 Event: Initial Sleep Onset Time Using Kleitman's DO Paradigm

Let us consider data from 5 individuals. Data are represented in a horizontal vector beginning at some arbitrary time-locked time. The first line indicates initial sleep onset as the time that each person dropped the spool an indicated with the letter D. On the next line, the sleep-tracker's determination of initial sleep onset is indicated with the letter O. Each x indicates the passage of one 30-second epoch, for example.

```
Subject 1:
xxxxxxxxxxxxxxxxDxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxOxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Subject 2:
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxDxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxOxxxxxxxxx
Subject 3:
xxxxxDxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxOxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Subject 4:
xxxxxxxxxxxxxxxxxxxxxxDxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxOxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Subject 5:
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxDxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxOxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

In the above example, the sleep-tracker would be determined within acceptable limits for subjects 1, 2, 4, and 5 and not fell outside the 25 second range for subject 3. Thus, the sleep-tracker agreed 80% of the time. The median precision was 6 seconds.

3.7.2 Process: REM Sleep Using PSG

Let us consider one hour of data for one individual. Data are represented in a horizontal vector beginning at some arbitrary synchronous time. The first line indicates the progression of sleep and wakefulness for each successive 30-second epoch. On the next line, a sleep-tracker's classification of REM or not REM is indicated. On the third line the epoch scoring is shown where "h" is hit, "m" is miss, "f" is false alarm, and "c" is correct rejection (i.e., not REM).

```
First 30 minutes:
Polysomnogram: nnnnnnnnnnnnnnnnnnnnnRRRRRRRRRRnnnnnnnnnnRRRRnnnnnnnnnnnnnnn
Sleep-tracker: nnnnnnnnnnnnnnnnRRRRnnnRRRRRRRRRRRnnnnnnnnRRRnnnnnnnnnnnnnnn
Epoch scoring: ccccccccccccccccffffcmmhhhhhhhhfffcccccccmhhhccccccccccccccc
Second 30 minutes:
Polysomnogram: nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnRRRRRRRRR
Sleep-tracker: nnnnnnnnnnnnnnnnnnnnnnnRnnnnnnnnnnnnnnnnnnnnnnnnnnRRRRRRRRRR
Epoch scoring: cccccccccccccccccccccccfccccccccccccccccccccccccccfhhhhhhhhh
```

This example shows that there were a total of 20 hits, 3 misses, 9 false alarms, and 88 correct rejections. Thus, for this data set for this device the sensitivity is hits/(hits + misses) = 20/(20+3) = 0.87 and specificity is correct rejections/(correct rejections + false alarms)= 88 / (88+9) = 0.91. The overall agreement rate is (hits + correct rejections)/(hits + misses + correct rejections +false alarms) = (20+88) / (20+3+9+88) = 90%.

Table 1. Performance Evaluation Approaches for Sleep Parameters

For the purpose of evaluation, parameters are categorized as (a) Events, (b) Processes, or (c) Patterns. Each parameter may be evaluated using a Direct and Indirect (Correlative) approach. Furthermore, some measures may involve more specific testing methods (paradigms). The twelve general approaches include (see abbreviation notation at table's bottom): DO; SR; DO with SR; Kleitman's DO paradigm; SRUA Paradigm; CT; Bed Sensor; PSG; EM Sensor; EEG Sensor; DLMO and ACT.

Parameter Name	Parameter Type	Direct Method	Correlative Method
TATS Start Time	Event	Self-Report (SR)	Lights-Out**
TATS End Time	Event	SR	Lights-On**
TIB Start Time	Event	Direct Observation (DO)	Bed Sensor
TIB End Time	Event	DO	Bed Sensor
Awake	Process	DO and SR*	PSG
Asleep	Process	DO or SRUA Paradigm	PSG
Awakening from Sleep	Event	DO with SR	PSG
Brief Awakening	Event	DO with SR	PSG
Initial Sleep Onset Time	Event	Kleitman's DO paradigm	PSG
Final Awakening Time	Event	SR or DO	PSG
Brief Moment of Sleep (Dozing)	Event	---	---
REM Sleep	Process	None	PSG
N1 Sleep	Process	None	PSG
N2 Sleep	Process	None	PSG
N3 Sleep	Process	None	PSG
CNS Arousal	Process	None	PSG
Dream Sleep	Process	SRUA paradigm	PSG or (EM &/or EEG sensor)
Core Sleep	Process	DO or SRUA Paradigm	PSG or (EM &/or EEG sensor)
Restless Sleep	Process	None	PSG or ACT
Sound Sleep	Process	None	PSG or EEG sensor
Circadian Amplitude	Pattern	Core Temperature (CT)	---
Circadian Period Length (tau)	Pattern	CT	---
Circadian Phase	Pattern	CT, DLMO	---

Notation:

SR= Self Report
DO= Direct Observation
PSG= Polysomnography
EEG= Electroencephalography
EM= Eye Movement
CT= Core Temperature
ACT= Validated Actigraph
SRUA= Self Report Upon Awakening
DLMO = Dim-Light Melatonin Onset

Notes:

* DO should be used to identify active wakefulness but SR is needed to determine quiescent wakefulness.

** Lights off and Lights on are terms used in sleep laboratories and indicate when the participant is instructed to go to sleep and wake up, respectively. However, in an *ad lib* environment when a person is free to sleep when he or she wishes, lights-out and lights-on do not necessarily indicate intention and attempt to sleep.

4 Compliance

The following section provides details concerning criteria for labeling a wearable or in-bedroom device as "compliant with the CTA standard". There are two categories of compliance established herein. The first is for "Sleep/Wake" determination and the second for "Sleep Stage" determination. The tester, after testing a device, may indicate the percentage agreement with essential sleep/wake measures and/or sleep staging measures. If the metrics derived from testing meet or exceed the "compliance" level as identified in 6.2, the manufacturer shall indicate that the device "Meets the CTA Standard".

4.1 Essential Sleep/Wake Measures [Mandatory]

1. TIB Start/End or TATS Start/End (Clock Times)
2. Awake/Asleep (Minutes)
3. Awakenings from Sleep (Count)

4.2 Essential Sleep Staging Measures [Optional]

If the tester chooses to trifurcate sleep into REM, Light (combined N1 and N2 Sleep), and N3 sleep, they must evaluate performance for:

a. REM Sleep and (combined N1 and N2 sleep) and N3 Sleep, or
b. REM Sleep and (combined N1 and N2 sleep) and Sound Sleep, or
c. REM Sleep and (Core minus Sound Sleep) and Sound Sleep, or
d. REM and (Core minus Sound Sleep) and N3 Sleep, or
e. Dream Sleep and (combined N1 and N2 Sleep) and N3 Sleep, or
f. Dream Sleep and (combined N1 and N2 Sleep) and Sound Sleep, or
g. Dream Sleep and (Core minus Sound Sleep) and Sound Sleep, or
h. Dream Sleep and (Core minus Sound Sleep) and N3 Sleep.

5 Epoch Alignment

The test device and the validated device used for comparison may employ different epoch lengths and/or have different epoch alignment. It is not possible to anticipate all situations, so the tester should choose a sensible and supportable approach to align the data for direct comparison. Clinical PSG uses 30-second epochs. To meet the present standard the comparison between devices must be minute-by-minute (or faster). Several examples are given below to illustrate reasonable options.

Example 1

Consider a test device being compared to PSG, in which the test device also uses 30-second epochs, and the epochs for the test device and PSG are automatically aligned with each other, e.g., by each being aligned with the same clock time. In this case the epochs from the test device and PSG may be compared directly.

Example 2

Consider a test device being compared to PSG, in which the test device uses 30-second epochs, but the epochs for the test device and PSG are not aligned with each other. For example, consider that the test device epochs are aligned to clock time, but the PSG epochs are aligned with the start of that recording. One sensible approach is to shift the data from the test device by the least amount to align the epoch start times with the PSG, i.e., shift the data from the test device to align with the PSG epoch with which it already had maximum overlap.

Example 3

Consider a test device being compared to PSG, in which the test device uses 1-minute epochs aligned with clock time, but the PSG recording uses 30-second epochs aligned with the start of that recording. PSG scoring already uses majority rule within 30 second epochs, i.e., each 30-second epoch is assigned a single sleep stage that represents the majority of that epoch. To accommodate the different epoch lengths and clock alignment, one sensible approach is to 1) report the test device directly, and 2) assign to each clock minute a PSG score taken as the PSG score that spans the majority of that minute. For

example, consider the case in which PSG epochs start at HH:MM:14, HH:MM:44, etc. The minute from HH:MM:00 to HH:MM:59 would be assigned the score from PSG epoch starting at HH:MM:14, because that epoch spans a full 30 seconds of that minute. The PSG and wearable scores in each clock minute may then be compared directly.

6 Sleep Classification

6.1 Nomenclature

The following formulae for detection statistics are expressed in the standard language of true positive (TP), false positive (FP), true negative (TN), and false negative (FN), where Sleep is the "Positive" class (*P*) and Wake is the "Negative" class (*N*). Consistent with the language of Section 3.7, TP = "hit", TN = "correct rejection", FP = "false alarm", and FN = "miss". As defined in Section 2.4, TD refers to the test device, and GS refers to the gold standard device.

6.2 Levels of Compliance with Standard

This standard supports two levels of compliance related to sleep/wake classification and sleep stage classification.

- Firstly, any device that meets the standard must classify sleep versus wake. This two-class comparison permits a variety of statistics. Overall accuracy (ACC) is easiest to interpret, however, ACC alone may be misleading, because data collected at night are likely to be dominated by sleep. For this reason, Sensitivity (TPR) and Specificity (1 – FPR) are more informative. Cohen's Kappa (κ) is a measure of inter-rater agreement that attempts to account for chance agreement and is used commonly in publications of inter-rater agreement and device performance. A device tested against this part of the standard must report all four of these statistics.
- Secondly, devices may optionally classify sleep stages. Each epoch is classified as either Wake, REM Sleep, Light Sleep, or Deep Sleep. This four-class comparison also permits a variety of statistics but, for simplicity, the comparison statistic here is limited to ACC. A device tested against this part of the standard must report ACC only.

6.3 Data Selection/Rejection

Before comparing, testers should review their data and discard any subjects (or nights) with obviously bad data. The report should describe and justify what data were acquired and selected for comparison, and justify discarding any data, as would be done in a scientific publication.

6.4 Sleep/Wake Classification (2-way)

Performance statistics for Sleep/Wake classification are based upon the following 2x2 confusion matrix.

		TD	
		Wake	Sleep
GS	Wake	*TN*	*FP*
	Sleep	*FN*	*TP*

Let *M* = the total number of epochs being compared = *TP* + *FP* + *TN* + *FN*.

Accuracy

Accuracy is defined as

$$ACC = \frac{1}{M}(TP + TN)$$

This quantity must be computed and reported in two ways. The first way (Within-Subject ACC) may be considered easier to understand and communicate. The second way (Across-Subject ACC) may be considered to have more statistical validity.

6.4.1 Within-Subject ACC

In the first approach, compute ACC for each subject/night, then compute the mean across subjects. For example, in a recording session, the percentage of 30-second epochs in which the TD agreed with the GS for 5 individuals might be 90%, 88%, 92%, 91%, and 85%, respectively. This yields an inter-subject mean of 89.2%. If multiple nights are recorded for an individual, the mean for that person should be calculated first and then the mean should be determined across subjects. This first approach to computing ACC is included because it is the simplest to understand and communicate, however, the average of averages loses some information so this not ideal for assessing compliance with the standard.

6.4.2 Across-Subject ACC

In the second approach, concatenate all the epochs from all subjects/nights, and record epoch-by-epoch the sleep/wake classification for the two devices. Note that concatenating epochs is equivalent to summing TP across subjects/nights, summing FP across subjects/nights, etc., then computing a single value of ACC.

6.4.3 Other Across-Subject Statistics

Because sleep data are typically dominated by the Sleep state, ACC alone may be misleading. The following statistics must also be computed in analogy with Section 6.4.2, i.e., concatenate epochs across subjects/nights, then compute a single value for each of the following statistics.

Sensitivity & Specificity

True Positive Rate (TPR) and False Positive Rate (FPR) are both normalized by the GS, i.e., not the TD. TPR should be interpreted this way: Of all the epochs for which the GS scored sleep, TPR is the fraction of epochs that the TD scored as sleep. FPR can be interpreted this way: Of all the epochs for which the GS scored wake, FPR is the fraction of epochs that the TD score sleep.

$$Sensitivity = TPR = \frac{TP}{TP + FN}$$

$$Specificity = 1 - FPR = 1 - \frac{FP}{FP + TN} = \frac{TN}{FP + TN}$$

Cohen's Kappa (κ)

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

where p_0 = ACC, and p_e = the estimated probability of chance agreement:

$$p_e = \frac{1}{M^2} \sum_{c=P}^{N} M_{1c} M_{2c}$$

Where M = the total number of epochs, M_{1c} = the total number of epochs that device 1 (i.e., the TD) scored as class c, and M_{2c} = the total number of epochs that device 2 (e.g., the GS) scored as class c. The sum runs over the two classes $c = P, N$ corresponding to Sleep and Wake, respectively.

6.5 Sleep Stage Classification (4-way)

The following statistics are computed in analogy with Section 6.4.2, i.e., concatenate epochs across subjects/nights, etc., and record epoch-by-epoch the sleep stage classification for the TD and the GS.

		TD			
		Wake	REM	Light	Deep
GS	Wake	M_{00}	M_{01}	M_{02}	M_{03}
	REM	M_{10}	M_{11}	M_{12}	M_{13}
	Light	M_{20}	M_{21}	M_{22}	M_{23}
	Deep	M_{30}	M_{31}	M_{32}	M_{33}

Here M_{12} is the total number of epochs the validated device scored as REM, and the test device scored as Light Sleep, etc.

Accuracy
Compute the overall accuracy ACC, i.e., the sum of diagonal elements in the table divided by M = the total number of epochs compared:

$$ACC = \frac{1}{M}\sum_{c=0}^{3} M_{cc}$$

where M_{cc} is the value along the diagonal corresponding to class c. Note that, by including the Wake class in this statistic, it is partly redundant with the required Sleep/Wake classification, however, this approach of including all 4 classes in *ACC* simplifies the calculation and interpretation.

7 Reporting

A Test Report shall be generated documenting the testing conditions and the test results of the sleep tracking consumer technology device/application.

Reporting shall be at a sufficient level of detail to allow a third-party evaluator to replicate the testing conditions and to produce similar test results.

Testing condition aspects:

1. The properties of the sleep tracking consumer technology device/application, including:
    a. Model number
    b. Update/revision number (if applicable)
    c. Firmware version number (if applicable)
    d. Location and/or position of the sleep tracking consumer technology device/application on the Participant during testing (if applicable)
2. Exclusion and/or inclusion criteria used for selecting Participants.
3. Description of the characteristics of each Participant including:
    a. Gender
    b. Age
    c. Health Status and relevant characteristics (e.g. physical conditions)
4. Detailed description of the testing conditions including:
    a. Equipment used
    b. Setting of testing equipment
    c. Sample Rate
    d. Statistics Used

The Test Report shall include test results for each of the sleep tracking consumer technology device/application as outlined in Section 6.2.

Consumer Technology Association Document Improvement Proposal

If in the review or use of this document a potential change is made evident for safety, health or technical reasons, please email your reason/rationale for the recommended change to standards@CTA.tech.

Consumer Technology Association
Technology & Standards Department
1919 S Eads Street, Arlington, VA 22202
FAX: (703) 907-7693 standards@CTA.tech


CTA™
Consumer Technology Association, producer of CES®