A CRITICAL REVIEW OF THE
"LEARN-WHILE-YOU-SLEEP" STUDIES

Charles W. Simon and William H. Emmons

P-534

14 June 1954
Revised 14 October 1954
A Critical Review of the "Learn-While-You-Sleep" Studies

Charles W. Simon and William H. Emmons

The RAND Corporation

Approximately 22 years of the average man's life is lost in sleeping. Economically minded persons and harassed students have long searched for some means to use this time to further advantage. For nearly a third of a century, now, there has been a growing interest in the possibility of trying to learn while one sleeps.

The science fiction writers were among the first to propose sleep-learning as an educational technique. In 1911, a magazine called "Modern Electrics" published a fiction story by Hugo Gernsback (4) in which the hero learned during sleep. The process was simple. Upon retiring, the student placed the material to be learned on a machine which played it automatically while he slept. In 1932, Aldous Huxley (10), in his Brave New World, scoffed at the possibility of improving the S's intellect while he slept, but described the world of the future as one which utilized the sleep period to train the lifelong attitudes of its populations. Other science fiction writers have continued to use the idea of sleep-training in their stories.

Popular news and picture magazines, citing experimental evidence, constantly reinforce the public's interest and misinformation about this osmotic form of education. From time to time, a national news service will release a story telling of some individual who works

* S refers to subject.
for a living by day and studies chemistry, operatic arias, or other lessons while sleeping at night. On "People are Funny," a popular radio stunt show, an audience participant learned a few short sentences and a girl's name while he was apparently asleep on the stage before the audience.

Following closely on the heels of public interest have come the entrepreneurs. Commercial firms have sprung up throughout the country selling recording devices which automatically turn on and off to permit the individual to learn while he sleeps. Impressive publications have been distributed by many of these firms propagandizing their products and extolling the validity and value of sleep-training (21, 23). On the whole, their claims of success for sleep-training have been extrapolations beyond the available supporting data. The majority of claims have been based on distorted facts, statements by unqualified authorities, and armchair hypotheses. Noncritical use of the information, anecdotal evidence, and the citing of inadequate research, much of which will be reviewed in the present paper, have made these commercial publications poor criteria for the support of the validity of the sleep-learning process.

In its constant search for new methods to speed training during national emergencies, the military, too, has tried to sleep-train its personnel. In both the first and second World War, exploratory efforts were made to teach service men Morse code (9,26) during their sleep. A Veterans' Administration Hospital doctor reported the successful use of sleep-training as
a supplement to waking-training in a mental health program (22). Unfortunately, most of these incidents have been too inadequately controlled to judge their effectiveness during sleep, and in general their results have not been sufficiently striking to justify the extra energy and expense required to carry them out.

It is the scientist who has been slow in investigating these claims for successful sleep-learning. Although psychologists have been concerned with problems of learning from their formalized beginning, there has been relatively little interest in the psychological processes which take place during sleep and a rather skeptical—although untested—attitude toward the value of studying sleep-learning. One of Pavlov's students, Krasnogorski (12), did attempt to condition the salivary reflex in a young child during sleep without success. Recently, however, a resurgence of interest in the problem has occurred. In the present paper the available research on the sleep-learning phenomenon has been collected and subjected to a critical evaluation.

Ten such studies were found. Of these only three have been published (B, F, G). Five were academic theses at the Master's or Bachelor's level (D, E, H, I, J) of which one was read at an APA meeting (I). The remaining two (A, C) were described in private communications from the authors. Although a number of other references were found, insufficient details were available for an adequate review or else they could not be classified as research under even the broadest definition.
The criticisms in this paper are not for the purpose of belittling these pioneer studies. These criticisms are believed necessary since these studies--many of which are unpublished--are being cited and their conclusions accepted both by the general public and some scientists. Because of the unusual methodological problems in this area of research, a review may help future workers to avoid similar mistakes.

Description of the Sleep-Learning Studies

The experiments reviewed in this paper are described briefly below.

In some of the studies, Ss were eliminated for various reasons. Only the number of Ss actually completing the experiment is reported.

A. Thurstone (26), in 1916, supplemented the waking training of 16 Navy men in a Morse code course with material presented during sleep. The criteria of sleep used in this study are unknown. The Ss finished the course three weeks earlier than had been expected. Thurstone concluded that the results "indicated some gain" for the sleep-trained group in their ability to send and receive Morse code.

B. LeShan (13), in 1942, tried to break the fingernail biting habit of 20 boy campers, ages 8 to 14, by playing the phrase, "My fingernails taste terribly bitter," through a loudspeaker 300 times a night for 54 nights. Ss were asked if they were awake before turning on the input; if the Ss appeared restless, the input was turned off. The nails of the Ss in the experimental group and an equal number of Ss in a control group were checked every two weeks for eight weeks for signs that the nail biting had stopped. Since 40 per cent of the experimental group and none of the control group stopped biting their nails, LeShan felt this indicated "the possible therapeutic use of suggestion during sleep."
C. LeShan (14), in 1943, taught a single S a different list of nonsense syllables each morning for 12 days. On the fifth and eighth nights, the list to be learned the next day was repeated 50 times while the S slept. The criteria of sleep used in this study are unknown. Fifty per cent fewer trials were required to learn the two sleep-trained lists than to learn the ten non-sleep-trained lists. LeShan concluded that the sleep-training facilitated learning.

D. Elliot (2), in 1947, studied the effects of 30 repetitions of a list of 15 common three-letter words during sleep on the per cent saved in learning the list by the anticipation method the following day. Two groups of 20 male college students of equal learning ability spent one night adapting to the laboratory environment. The word list was repeated serially to one group of Ss while they slept. An EEG record was used as a criterion of sleep. No material was played when the S showed "clear alpha patterns" although the machine was turned off when the E* believed the S would stay asleep. On awakening the next morning, the two groups learned the sleep-trained list by the anticipation method and were asked if they awoke during the night. The group receiving the sleep-training showed/significantly greater per cent of savings than the control group (P = .05)². No significant differences between groups were found on the basis of errors or absolute number of trials to learn. Elliot concluded that there is some retention of auditory material during sleep.

E. Hedges (5), in 1950, tried to improve the speech of three mentally retarded and aphasic children (ages 11, 7½, and 6) by sleep-learning. Short sentences, paragraphs, and simple consonant sounds were played to the children.

* E refers to experimenter.
between 433 and 1501 times distributed over 7 to 13 nights. The machine was not turned off if a S awoke. The third child required fewer waking lessons to learn the sleep-trained consonant than it had required previously to learn a non-sleep-trained consonant. The babbling of the first child increased. The second child showed no effects. Hedges interpreted this as showing "the possibility of perhaps speeding" the training of the third S.

F. Fox and Robbins (3), in 1952, taught 30 college males and females a list of 25 pairs of English-Chinese vocabulary. Ss were divided into three groups, matched on a pretest. Two vocabulary lists were prepared; in the first list, word pairs were matched as they were to be learned on the posttest (the facilitation list), while in the second, they were mismatched (the interference list). These lists were played 15 times to two of the S groups approximately three hours after they had retired to go to sleep. The third group—a control—heard only music while they slept. Although the Es did not observe the Ss during the input period, they disqualified all Ss who said they had awakened during the night. The facilitation group required fewer trials and the interference group required more trials to learn the post training list than did the control (P < .001). The Es concluded that learning could occur during sleep and could be detected by the savings method.

G. Leuba and Bateman (15), in 1952, believed that they had taught the words of three songs, 8 to 27 lines long, to a lightly sleeping S. Each song, unknown to the S beforehand, was played five times a night for three successive nights to the sleeping S. During the input period, the S was not observed by the E, although for three brief periods, the S's husband observed no restlessness on the part of the S. The S claimed she did not awaken during
the night. Given only the song title the next day, the S was able to write
the lyrics of two out of the three songs with no errors and of one with only
minor errors. The Es believed the sleep-training was successful. No learning
occurred in a later study during and following the use of sedatives.

H. Hoyt (8), in 1953, taught ten pairs of English-Chinese vocabulary
to his Ss. Twenty Ss were matched on a pretest and given one night to
acclimate to sleeping in the laboratory. On the experimental night, one
group of eight Ss received twelve repetitions of a list paired in the same
manner as that to be learned the next morning (facilitation list). A second
group of eight Ss received a list of comparable words, but with the pairs
mismatched (interference list). The remaining four Ss acted as a control
group and received no sleep-training. During the input period, Ss were ob-
served by E. If they awoke or heard the stimulus material, they were to turn
on a light. If this occurred, or if they stirred, the recording was turned
off until the S lay quiet for a suitable length of time. The next morning
Ss were asked if they had awakened during the night. Those that had not
were given the paired vocabulary list to learn to two correct anticipations.
Statistically insignificant differences were found to favor greater savings
for the interference group rather than the facilitation group in both the
mean number of trials required to learn the list as well as the number of
correct responses occurring before the criterion was reached. No comparisons
were made with the control group. Hoyt felt that under the conditions of
this study, learning during sleep could not be detected.

I. Stampfl (25), in 1953, had six college males learn lists of ten non-
sense syllables while acting as their own controls. Different lists were
repeated 4, 8, 16, and 32 times on different nights while the Ss slept, and were learned to one correct antitipation on the morning following the repetition of a particular list. At other times, the same Ss were tested on other lists after no sleep-training. Before presenting the stimulus material, Ss were asked if they were awake. During the input period they were watched, and if movements occurred, the input was stopped. Since no significant differences were found for savings in either trials or errors between learning sleep-trained and non-sleep-trained lists, no comparisons were made among performances after different numbers of training repetitions. Stampfl felt that the sleep-learning hypothesis was uncertain and improbable.

J. Coyne (1), in 1953, carried out a series of exploratory studies on a variety of psychological problems using from four to six male college students as Ss. He believed the results to be generally favorable for sleep-learning. Unfortunately, a statistical error was discovered in Coyne's thesis after he had written his discussion and conclusions which lead to a more favorable interpretation of the results than was justified. Although the E could observe the S as he slept, he primarily depended on the S to press a buzzer as soon as he awakened during the night. The problems Coyne studied are described below.

1. An interference list of adjectives was presented 25 times to the sleeping S. A similar list was learned to one perfect antitipation the next morning. A control group had no sleep-training. The sleep-trained group did poorer than the control group (P = .07).

2. When the first problem was repeated using the S as his own control, more errors were made on the list which had received the interfering
sleep-training (P = .15). No difference in savings (trials to learn the post-training list) was observed.

3. Twenty pairs of numbers and words were presented 24 times during a single 45 minute period to the sleeping Ss. These same Ss also received another list for the same number of repetitions distributed over a four hour period. On subsequent mornings, Ss were required to answer the appropriate word when given the number found on the list trained during the night, after that list had been mixed with ten additional new number-word pairs. A greater percentage of sleep-trained words were associated with the correct numbers than those words not trained during sleep (P = .15). Distributed sleep-learning resulted in fewer recall errors than did massed sleep-learning (P = .05).

4. While the Ss slept, one list of adjectives was repeated 100 times; on a different night, another list was repeated 25 times. On the morning following the sleep-training, the list was learned to one perfect repetition. No significant differences were found in performance between the two amounts of training.

5. Ss were required to solve a number of problems by finding the E's solution of a concept composed of the correct combination of letters, colors, and their relative positions. The solution to one problem had been repeated 180 times while the S slept the night before. Insignificant differences were found favoring the performance on the sleep-trained problem.

6. While Ss slept, one list of nonsense syllables was presented 30 times beginning at two o'clock and, on a different night, another list was presented 30 times beginning at five o'clock. Since it was discovered
that the two lists were of unequal difficulty, comparisons between performance at different times of presentation were not made.

7. During sleep, the contents of a particular picture were described to the Ss 90 times. The following morning, a number of out-of-focus pictures were shown, including the one described during sleep. The degree of focusing required before the S could identify the pictures was the measure of learning. No significant differences in the case of identifying sleep-trained and non-sleep-trained pictures were found either when Ss were asked to give unaided responses or when multiple choice solutions were provided.

Rather than examine each of the ten sleep-learning studies independently, this paper has been organized to discuss them all on the basis of the following categories: experimental design, statistical considerations, methodological considerations, and sleep criteria.

Experimental Design

The use of a control group or of using the S as his own control is recognized as a necessary procedure in order to know whether a certain experimental effect is real or not. In a number of the sleep-learning studies, however, both of these techniques were conspicuously absent or inadequately handled.

Thurstone (A) recognized the inadequacy of his uncontrolled experiment teaching navy men Morse code and attempted to run a second study in order to compare the performance of one group which received sleep-training with another which did not. This study was discontinued,
he reported, when it was discovered that ambitious navy instructors of the control group had been giving extra daytime instruction in order not to be out-taught.

Hedges (E) used no controls with two of the three speech-defective children he attempted to sleep-train. Since improvement in speech is a maturational problem, even for retarded children, the need for a control was paramount. The increased babbling of the first child may have been independent of the sleep-training and due solely to an additional month or two of growth, although Hedges believed that it took place immediately after the introduction of the sleep-training. Since the S also received waking-training on the same material, it is impossible to know to what extent one can attribute the increase in babbling to sleep-training. No apparent learning took place with the second S. Although Hedges' third S acted as his own control by learning to pronounce one consonant with sleep-training added to his waking-training and another consonant without the sleep-training, this was an inadequate control measure since the experimental design was such that the sleep-trained consonant followed the non-sleep-trained consonant. Any improvement in the latter could be attributed to practice as well as maturation.

Nor was a control used in the work of Leuba and Bateman (G). In this study, the S presumably had no previous knowledge of the songs played to her during sleep, yet was able to write the lyrics without further training on awakening. If this were the case, any
learning which took place during sleep would be a significant improvement, although materials such as songs and poetry probably have some internal predictability. This form of experimental design represents an implied control. It is implied that the S acted as her own control for had she been tested previously, no appropriate responses could have been made, nor could one suspect that maturational factors were operating to produce positive results.

A number of Es actually used their Ss as their own control. In LeShan's (C) second study, he compared the number of trials a S required to learn a list with and without sleep-training over a period of days. By having the non-trained periods before and after the sleep-trained periods, the superiority of the sleep-trained lists could not be attributed simply to practice or maturation. On many of his studies, Coyne (J) gave sleep-training to his Ss one night and no sleep-training on the next, counterbalancing the order for these procedures between two groups of Ss. However, when the S acts as his own control either the study material which is used under the varying conditions must be carefully equated, or additional counterbalancing must be introduced into the experimental design to correct for the inequalities. No E used this counterbalanced design; some made use of the published tables on which similarities and associability of the stimulus material had been previously calculated for the items. Several of Coyne's (J) studies were unanalyzable after he discovered the study material had not been equated.
Four of the experiments were designed to use separate control groups. In Elliott's (D), Fox and Robbins' (F), and Hoyt's (H) studies pretesting took place in order to equate the mean performance of the control and experimental groups. LeShan (B) used unmatched groups in his study with fingernail-biting children. The median age of his twenty experimental nailbiters was slightly less than ten years. His control was divided into two groups consisting of eight nailbiters with a median age of nine years, and twelve more nailbiters with a median age of twelve. Some question might be raised concerning LeShan's failure to better equate the experimental and control groups on age, for there is reason to suspect a relation between nailbiting and age. Wechsler (28) found a sharp rise in nailbiting for boys around the age of twelve; if this is so, there would be a lower probability for the older control group to stop biting its nails than the younger experimental group, thus reducing the effectiveness of the older group as a control.

Two of the experimenters added a second experimental group to their design. Fox and Robbins (F) and Hoyt (H) used both a facilitation and an inhibition group in order to obtain a more sensitive indication of the value of sleep-training. Of all of the studies, only that of Fox and Robbins (F) provided the control group with a neutral stimulus—music—for the same amount of time that the experimental group received the verbal test material. If there are any disturbances during sleep due to the stimulus and if these in turn affect recall,
such a procedure is a wise one.

Although none of the Es were directly concerned with the problem, the use of an additional control group to compare results from sleep-training with the equivalent amount of waking-training would have been quite illuminating.

Statistical Considerations

Only five of the Es (D, F, H, I, J) treated their data statistically to see if the sleep-training improved performance significantly. The remaining Es used clinical criteria to evaluate the effects of sleep-training.

Although Elliott's (D) results favored the performance of the sleep-trained group over the non-sleep-trained group, he failed to find a significant difference at the 5% level in the number of trials it took to relearn the training list. Elliott had equated his groups on a pretest, but did not attempt to match the individual Ss for the analysis. Since equating tends to decrease differences between means, failure to remove the variance due to Ss inflates the error variance and decreases the probability of getting significant differences. When the present authors (24) did an analysis of covariance with pre- and posttest scores from Elliott's data, the differences between the mean number of trials to learn a new list by sleep-trained and non-sleep-trained Ss were significant below the .05 probability level.
Coyne (J) used the one-tailed t test to evaluate his data. The wisdom of this treatment is questionable. In order to avoid abuses and controversy, exploratory work should be as cautious in its interpretations of results as it should be daring in attempting new ideas. Using the one-tailed t test does not allow for the possibility that differences might be in the direction opposite to that hypothesized (6). This is serious in any exploratory work; it is particularly dangerous in sleep-learning studies where one could seriously suspect that the intervening training during sleep might actually hinder normal waking recall.

Half of the Es (A, B, D, F, H) used a reasonably large number of Ss as compared to the number used in most psychological studies; in the remaining cases, the number was smaller. When the number of Ss is small, one might be more critical of accepting the null hypothesis merely because the level of significance was not below the traditional 5% level. For studies as exploratory as these, significance levels of 15% could be arbitrarily considered encouraging. Since the expense in time and money is relatively small, preliminary work in sleep-training should favor fewer Type II errors in order not to reject valuable experimental leads by accepting a false null hypothesis.
Methodological Considerations

A number of Es believed that differences in methodology might have been responsible for the divergent successes and failures found in the sleep-learning studies. These and other considerations are discussed below.

Subjects

The majority of the studies employed the traditional college student—male and female—as Ss. Thurstone (A) used navy men and LeShan (B) used young boy campers. Hedges (E) bravely attempted to provide sleep-training as a supplement to the waking-training of children who had speech deficiencies and who were suspected of being mentally retarded. None of the Es attempted to study the effects of either age or sex on sleep-learning.

The selection of Ss may have an effect on whether successful sleep-training results are attained or not. Underwood, while reviewing Fox and Robbins' paper in the 1953 Annual Review of Psychology, commented that "such low variability in performance scores among Ss on the test list is rarely found in normal transfer experiments with such material, but this may again only reflect the presence of a highly select and homogeneous sample" (27, p. 48). Low within-group variability was certainly in part responsible for the high degree of significance of the differences between the sleep-trained groups and the control.
Whatever the variability of the group, it would appear wiser to choose individuals who have shown the capacity to learn while awake. Perhaps the effects of sleep-training are so subtle that its benefits will be found only when it is applied to individuals with very high IQs.

Number of Repetitions

The experiments can be divided into two groups on the basis of this variable—those who gave an exceptionally large number of repetitions of the material during sleep (E, H) and those that gave significantly fewer repetitions (C, D, F, G, I, J). The large number of repetitions were actually spread over a number of nights and ranged from a total of 433 times over a period of 8 nights to 16,200 times over a period of 54 nights. The smaller number of repetitions ranged from 8 to 180 times, with one study (G) playing the material 5 times per night for 3 successive nights. The more repetitions were characteristic of the field as opposed to the laboratory studies. Both groups obtained both positive and negative results, although the tasks were varied sufficiently so that direct comparisons could not be made. LeShan's (E) study, in which 40% of the experimental group stopped biting its nails, represented a successful example where a great deal of repetitious training seemed to have affected a semi-involuntary behavior.

Stampfl (I) and Coyne (J) found no differences when they tried to study the effects of different amounts of repetitions on sleep-learning. Actually since neither E used a non-sleep-trained group as a control, neither could conclude that any sleep-learning took place
at all in this phase of their study. Coyne (J) suggested that in his study there may have been no differences in savings after one hour of repetitions than after four hours because the material was simple enough to be learned in one hour and additional practice could not improve it. This is a reasonable hypothesis, as is its antithesis—that during sleep, learning is sufficiently slow so that little is learned even after four hours of repetitions. Sleep-learning, if it is to occur at all, may require that an extremely large number of stimulus repetitions take place. It will be important to evaluate sleep-versus waking-training from the standpoint of economy of both time and effort.

Presentation

The manner in which the study material is presented to the sleeping S has been considered by some Es as critical to the success or failure of the training. Two dimensions of this variable are the time of presentation and the order of presentation.

The problem of presentation time is an important one since during the sleeping period, time is related to some extent to the depth of sleep, which in turn may be related to trainability. Deeper sleep tends to be more prevalent in the early period of sleep, while lighter sleep tends to occur later (20). Of course, the levels vary considerably throughout a normal night's sleep.
Coyne (J), studying the effects of presenting the material at different times during the sleeping period, failed to equate his lists beforehand and could draw no conclusions from his results. He recognized that presentation time might be inversely related to the amount retained. He suggested, however, that this was due to the recency of the presentation to the recall period. It is interesting to speculate whether or not the Jenkins-Dallenbach interference effect (11) occurs within the sleep period for materials presented during sleep.

The differences Coyne found between massed and distributed learning might also be accounted for on the basis of presentation time. Although he concluded that distributed sleep-learning was superior to massed sleep-learning, there is no way to determine whether this was so because of the spacing of the training or because some of the distributed inputs occurred during the period just before waking—often a light drowsy state—while the massed inputs occurred only during the deeper and possibly less receptive period.

The order in which the material was presented during sleep may also affect the results of sleep-training.

In a waking state, serial learning has been shown to be easier than learning material presented in a varying order (7). Thus, if any learning takes place in sleep, a serial presentation would more probably increase any positive effects which might occur. Also, if the sleep state is one in which no mental organization takes place, this would favor the learning of only the more organized serial presentation.
Hoyt (H) and Stampfl (J) varied the order in which their material was presented and failed to find that any sleep-learning took place. Fox (F), LeShan (B), Elliott (D), and Leuba (G) believed they found evidence of successful sleep-learning. Thurstone (A) probably varied the order of his material, but his Ss were practiced over a period of months so that the effects of presenting the material in a varying or an unvarying order may have been minimized.

The varying presentation order in Hoyt's study was methodologically different in one major respect from that used in Stampfl's study. Since the former study used paired associate material, varying the order in which the pairs were presented would have less effect than it had in the latter study where the order of words in a list was varied during sleep-training, even though they had to be recalled serially during the waking period. Any positive effects of sleep-training in the latter case may have been neutralized by the negative effects built up through nonserial learning.

Training Problems, Materials, and Mode of Input

The types of psychological problems studied by the majority of sleep-training investigators have been quite limited. With the exception of Coyne's (J) work on concept formation and perceptual sets and LeShan's (B) and Hedges' (E) therapeutic studies, the remainder of the research has been involved with training problems which require the memorization of word lists. It is unlikely that all types of problems are suitable for sleep-training, although exploratory studies should examine many rather than a few possibilities.
Of the material used in the studies requiring verbatim recall, the degree of meaningfulness ranged from lists of nonsense syllables through short words to foreign language vocabulary. We know that the more meaningful the material, the easier it is to learn in the waking state (7). Stampfl (I) suggested this might explain why his results with nonsense syllables were poorer than Elliott's (D) who used a list of adjectives, and why Fox (F), using a Chinese-English vocabulary, got even more striking results. This does not seem to be a critical variable, however, for Hoyt (H), using the same Chinese-English vocabulary as Fox (F), got negative results, while LeShan (C) and Coyne (J), using nonsense syllables, got positive results.

All of the Es used an auditory input. This is certainly the most obvious technique and would appear to be the most economical; however, other sensory channels need not be ignored. It may be necessary to flash lights on closed eyelids or to apply tactual stimuli to the fingertips in code in order to "reach" the sleeping S.

Techniques and Measures of Retention

The techniques used to measure the retention of training material can have considerable influence on the amount of material recalled. However, in the present studies, there did not appear to be a pattern of successes or failures consistent with the technique used.

A number of Es did not require the verbatim recall of the verbal material presented during sleep. Instead, they evaluated retention on the S's ability to use the material in posttraining tasks or on an
observed change in the S's behavior after the training. LeShan (E) examined his Ss' fingernails every two weeks to see if the children had stopped biting their nails. This technique might have been slightly more objective had the examiner not known which children were and were not receiving the sleep-training. Hedges' (E) clinical evaluations of his children's speech improvement required even more subjective judgments. Because of the very complexity of this measure, Hedges wondered whether it was sensitive enough to detect improvement. Coyne (J), in several studies, had his Ss describe an out of focus picture, the identity of which had been given to them during sleep, and to determine the E's solution of concepts composed of the correct combination of certain stimulus variables, the answer to which had also been provided during sleep. Thurstone's (A) Ss were evaluated on their ability to send and receive Morse code. Both Coyne's and Thurstone's results could be quantified.

Leuba (G) gave his Ss the title of songs played during sleep and required them to recall unaided the lyrics. Positive results were claimed for sleep-training in this study.

The savings method has been used by a majority of the Es (C, F, H, I, J), for they believed it to be a more sensitive measure of retention. Stampf1 (I) believed that although material could not be consciously recalled at any moment its presentation during sleep may have still modified the nervous system sufficiently to make learning easier and a savings effected. However, use of the savings technique
may confuse the measure of retention with a measure of the ability to learn since both are confabulated within the same performance score. None of the Es using the savings measure compared the time to relearn a list with the time it took the same S to learn an equivalent list presented after sleep. Comparisons with the original list presented before sleep are not sufficient. The positive transfer which occurs from the pretest to the posttest—the phenomenon of "learning how to learn"—may account for some of the apparent savings which several researchers believed they found. However, Stampfl (I) gave three practice lists to be learned before the experiment in order to bring the Ss' learning curve closer to its asymptote and thus reduce the "learning to learn" effect.

A still more sensitive technique for measuring retention is that of recognition; at least in Inh's (16) classical study this was so after a two day interval between training and the posttraining test. Coyne (J) used this technique in the form of a multiple choice test for one of his exploratory studies on perceptual set, but still failed to get positive results. Hoyt (H) told of an exploratory study in which he presented a single number-word pair to a sleeping S who was under constant observation for a total of 11 hours on three successive nights. On awakening after the third night the S was given the number and asked to pick the correct word from a group of ten.
He was unable to do this correctly. A major difficulty with the recognition method is that it must be corrected for chance guessing. With words, however, this correction for chance is difficult since all words do not have an equal probability of being recognized (19).

Variations in the score used to measure retention have failed to consistently differentiate success and failure in sleep-learning. Many of the Es (C, D, F, H, I, J) used the number of trials to learn posttraining material to one perfect repetition as a measure of retention. Hoyt (H) required that the lists be learned to two perfect anticipations; this may tend to make his results slightly more reliable. Some Es (D, H, I, J) also used as measures of retention the number of errors or correct responses made on the first trial or later trials, or on cumulative trials. Where the S's performance was used to measure retention, the measure was characteristic of the task, e.g. Coyne used the extent of focus of the projector as the measure of performance in his perceptual set study. The measure that is used may determine in part the results which are found. For example, Coyne (J) noted that the measure by trials often favored the opposite results than those favored by the measure by errors. Error measures generally lead to less conclusive results than those measured by trials. This situation does occur in waking-learning research and need not be contradictory, although it certainly affects the conclusions drawn as well as the practical applications of the results.
Before final conclusions can be drawn concerning the feasibility of sleep-learning, more recall techniques should be tried. Simon and Emmons (24) discuss this while considering the possibility of secondary cerebral storage mechanisms for material introduced during sleep\(^3\).

Sleep Criteria

Perhaps the most damaging criticism of the sleep-learning studies to date have been the inadequate control of sleep and the criteria used for defining sleep. So elusive is the process of sleep from the psychological standpoint and so little is known about the actual physiological mechanisms which cause sleep that the problem of determining whether a S is asleep is to a great extent a semantic one. Although both direct observation and the S's subjective report are reasonably reliable for deciding if a S has been asleep over a block of time, neither can be considered sufficient to know if a subject is asleep at any moment in time. Therefore, the care with which the E determines the sleep condition of his S at the time of the input is highly important and determines the degree of confidence which one may place in the conclusion that learning during real sleep is or is not possible. Some Es (D, F, J) eliminated the S from the experiment if he awoke or said he heard the stimulus. Other Es (B, I) shut off the machine until the S went back to sleep.

In four of the studies (D, F, G, H), the Es asked the Ss after the training if they had awakened. This is an unsatisfactory criterion of sleep for experimental studies since it is a common experience
to awaken during the night, perform a number of rational acts before
retiring again, and remember nothing the next day (8, 20). Hoyt stated that one S was observed to awaken during the night, leave the laboratory to go to the bathroom, and later return to bed. The next morning, the S remembered nothing about it. In two of these studies (F, G), the Es were not even present to observe the Ss as an added check.

In one experiment (F) the time of night was considered a partial criterion of sleep; training was applied during the period when studies have shown sleep to be at its deepest. However, published sleep curves are based on statistical averages and cannot be used to accurately predict the sleep characteristics of the single individual. Individual sleep curves vary considerably from time to time during the night (17).

Two of the Es (B, I), before presenting the stimulus material, asked the Ss if they were asleep. Since LeShan's (B) Ss were adolescent boys in camp, naive as to the purpose of such questioning, one may wonder if asking them if they were asleep would be a sufficient check. It has been observed that even willing Ss often fail to respond to such questions, although they admit hearing them later. Another major difficulty with this technique is that the S may actually have been asleep when the test question was asked, but later would awaken sufficiently to hear the subsequent input. In these studies, no systematic check was made to be sure that Ss did not awaken after the
initial test, except to note if they moved.

Coyne (J) and Hoyt (H) had their Ss press a button whenever they awoke in order to stop the input of training material. This method is not sufficient alone since it has been found that often the desire as well as the ability to do this is not present although Ss are able to consciously hear and understand the material. Since much of the material used in these studies was short and specific, even a few seconds of waking presentation would have been sufficient to allow Ss to hear and learn much of the stimulus while awake.

Stampfl (I), LeShan (E), and Hoyt (H), in addition to asking their Ss if they were or had been awake, also noted whenever they moved, and turned off the machine at that time. This criterion of sleep is not completely satisfactory since EEG records indicate that sleep may lighten to almost a waking state before movement occurs, as well as afterwards, or when there has been no movement at all (17). Hoyt (H) found that two of his Ss said they had awakened but showed no observable movements, while three subjects said they had not awakened but showed from two to six movements each. This suggests that the criteria of asking and noting movements did not correlate very well. In Hedges (E) study, the parents occasionally observed the children to see whether or not they awoke during the sleep-training; two children awoke. In the first case, the number of training repetitions were not counted for the time the S was awake. However, without a waking control, this correction is meaningless. The other S awoke while the
recording was playing and began drowsily to follow the instructions given on the record. Hedges felt this aided the progress observed in that case.

Elliott (D) used the electroencephalogram to determine the S's sleep level. Independent workers using a variety of techniques have established a significant correlation between the brain wave patterns and the depth of sleep (16). Because the writers believe that properly evaluated electroencephalograms represent the most objective, continuous, and practical on the spot indicator of sleep depth available, Elliott's (D) positive results—often quoted by the commercial firms in support of their claims for sleep-learning—are quite noteworthy. An examination of his sleep criterion revealed one major flaw, i.e. Elliott did not keep the EEG running during the entire training period. Therefore, no continuous check was made of the S's sleep level while the stimulus was being played. The machine was turned on at the beginning of the training and kept on until the E believed that sleep would remain deep enough; then he would shut it off for the remainder of the half-hour stimulus period. Using Elliott's data, a correlation of -.39 was found between the amount learned and the amount of monitoring. This was not significant at the 5% level but in a direction which suggested that more savings occurred with less monitoring. The writers (24) also compared the amount of savings made by a group of Elliott's five Ss who were monitored nearly all of the training time with the savings of the remaining fifteen Ss who were monitored on an average
slightly more than a fourth of the time, and never more than 54% of the time. For the first group, the average savings in trials to learn was 27.5%; for the second group, the average savings was 45.0%. The difference of 17.5% was highly significant below the 2% level of significance and we can conclude that when a thorough check of the sleep condition was made with the EEG during training, considerably less savings took place.

That any savings occurred for the groups monitored nearly 100% of the time might be accounted for in two ways without assuming that learning actually occurred during sleep. First, the apparent savings may be simply an effect of the "learning how to learn" phenomenon discussed earlier in the section on "Methodological Considerations". Second, Elliott stated that whenever an alpha pattern was "clearly" present he concluded the S was awake and turned off the input machine. In so doing, he quite likely permitted many of his Ss to listen while they were already awake since movement, tenseness, or the opening of the eyes may sometimes block a clearly established alpha.

The expense of the EEG equipment negates its widespread use. This does not mean that without it, however, adequate sleep-learning studies cannot be done. The use of a combination of the criteria and continuous monitoring may be sufficient to insure some rough but adequate control of the S's sleeping condition. It is interesting to note that Hoyt (H) and Stampfl (I), finding negative results, were the two Es using the greatest number of multiple and continuous criteria.
The continuity of monitoring cannot be overemphasized; for in
spite of wishful thinking to the contrary, Ss do awaken while the input
is being presented. Two out of three of Hedges' (E) Ss awoke. As many
as half of Coyne's (J) Ss awoke in some of his studies. Elliott (D)
reported that six of his Ss awoke. Only two out of sixteen of Hoyt's
(H) Ss in the experimental group failed to awaken during an input peri-
od. LeShan (B) also stated his Ss awoke, but did not specify how many.
One-fourth of Fox's (F) Ss awoke, and were eliminated from the study.

The conditions under which the Ss sleep can influence to some
extent the soundness of their sleep. Elliott (D) and Hoyt (H) required
the Ss to sleep in the laboratory one night previous to the experiment.
Any sleeplessness occurring on this first night probably induced a
dereper sleep on the second. Hedges (E), Fox (F), and Leuba (G) used
Ss while they slept in their own homes, while Thurstone (A), LeShan (B),
and Stampfli (I) used Ss who had been sleeping in familiar "homes away
from home" for some time.

Although most of this discussion has centered around the problem—
was the sleep state deep enough?—Stampfli (I) and Leuba (G) suggested
that there is an intermediate point between waking and deep sleep that
is optimal for sleep-learning. Some negative results, they believed,
may have been obtained because the S was too deeply asleep. Whether
this is true or not is an experimental question for future Es to answer.
Summary and Conclusions

Ten sleep-learning studies were reviewed. Many of these have been cited uncritically by commercial firms or in popular magazine and news articles as evidence in support of the feasibility of learning during sleep. A critical analysis was made of their experimental design, statistics, methodology, and criteria of sleep.

It is highly speculative whether or not the studies reviewed in this paper have presented any acceptable evidence that learning during sleep is possible. The inadequate control of a number of experimental variables makes the validity of the conclusions drawn by many of the Ss unwarranted. The conditions under which the results were found tend to support the contention that learning takes place in a special kind of waking state wherein Ss do not apparently remember later on if they had been awake. This may be of great practical importance from the standpoint of economy in study time, but it cannot be construed as sleep-learning. More carefully controlled experiments in the future may provide us with a clearer answer to the question: can one learn during sleep, as well as to provide comparative data between waking and resting learning from the standpoint of economy of time and effort. The problem is partially confounded by an inadequate definition of sleep.
References


14. LeShan, L. Personal communication. 27 October 1953.


Footnotes

1/ Capital letters in parentheses are used to identify the sleep-learning studies being reviewed.

2/ The P values representing the results of tests of significance are all reported as if they were based on a two-tailed t test, irrespective of what the author used. P levels greater than .15 were not reported.

3/ Dr. Bernard Fox, in a private communication (26 June 1954), described his attempt to use hypnosis as a means of facilitating the recall of sleep-trained material. A 55 year-old man was brought to a point where deep hypnosis was possible. In his normal sleep, he was presented with a list of ten words repeated for a half-hour. During this period, the S was observed and the input turned off whenever he moved. The next morning, he was unable to recall any words unaided. When he was hypnotized, he made only one association which might have been related to one of the words. Still under hypnosis, he recognized fewer words than would be expected by chance when the ten were interspersed randomly in a group of thirty. Neither reading the words aloud nor making him choose by a forced choice method produced any more positive results.

4/ Fox stated "if any Ss in the group of 30 actually did wake up without reporting it, then the number was probably distributed about equally throughout the groups. Moreover, since all Ss with one exception show effects in the expected directions, it is clear that the results cannot be explained on the bases of a few Ss in each group who awakened but failed to report doing so" (3, p. 78). It may be on the basis of many.