

RELIABILITY OF A SCALE ASSESSING DEPRESSED MOOD IN THE CONTEXT OF SLEEP

BRANDY M. ROANE

UNIVERSITY OF NORTH TEXAS HEALTH SCIENCE CENTER — BROWN UNIVERSITY
OF PROVIDENCE

RONALD SEIFER

BROWN UNIVERSITY OF PROVIDENCE — E.P. BRADLEY HOSPITAL OF PROVIDENCE

KATHERINE M. SHARKEY

BROWN UNIVERSITY OF PROVIDENCE

ELIZA VAN REEN

BROWN UNIVERSITY OF PROVIDENCE — E.P. BRADLEY HOSPITAL OF PROVIDENCE

TAMARA L. Y. BOND

BROWN UNIVERSITY OF PROVIDENCE

TIFENN RAFFRAY

BROWN UNIVERSITY OF PROVIDENCE — LAUSANNE UNIVERSITY HOSPITAL

MARY A. CARSKADON

BROWN UNIVERSITY OF PROVIDENCE — E.P. BRADLEY HOSPITAL OF PROVIDENCE

The current study assessed the reliability of Kandel & Davies (1982) mood scale with and without sleep-related items. Brown University first-year students (mean age = 18.1 years; 108 females) completed online biweekly surveys after weeks 2, 6, 8, and 10 and on two consecutive days after weeks 4 and 12 of their first semester. One hundred seventy-eight students completed at least two biweekly online surveys with 128 students completing all six surveys. The scale was examined as a 1) full 6-item scale, 2) 5-item scale excluding the sleep item, and 3) 4-item scale excluding the sleep and tired items. Intraclass correlations (ICC) values for consecutive-day assessments and six biweekly surveys were similar and not a function of the weeks evaluated. Total-item correlations and inter-measure correlations with the Center for Epidemiologic Studies — Depressed Mood Scale (CES-D) supported the removal of the sleep-related items from the 6-item scale. These analyses confirm the reliability of the original Kandel and Davies depressed mood scale as well as without the sleep-related items.

Key words: Sleep; Depressed mood scale; Reliability; Insomnia; Mood.

Correspondence concerning this article should be addressed to Brandy M. Roane, CBSM, University of North Texas Health Science Center, Department of Internal Medicine, 855 Montgomery Street, Fort Worth, TX 76107, USA. E-mail: Brandy.Roane@unthsc.edu

INTRODUCTION¹

The intricate connection between depressed mood and disturbed sleep presents a conundrum for researchers and clinicians interested in assessing depressed mood along with sleep. One model of the interplay between depression and sleep (Pigeon & Perlis, 2007) indicates that disturbed sleep can precede, follow, or co-occur with depressed mood (Manber & Chambers, 2009; Taylor, 2008). Inclusion of sleep items in standard depressed mood assessments reflects this entanglement of the two symptom clusters. For example, two of the 21 items on the Beck Depression Inventory-II (Beck, Steer, & Brown, 1996) and one of the 20 items on the Center for Epidemiologic Studies of Depression Scale (Radloff, 1977) are sleep-related. Furthermore, clinician-administered mood measures, such as the Hamilton Depression Rating Scale (Hamilton, 1960), include sleep symptoms. Finally, sleep disturbance is a cardinal symptom of major depressive disorder as defined by DSM-IV-TR (American Psychiatric Association, 2000), and mood disturbance is a daytime symptom of insomnia as defined by the International Classification of Sleep Disorders, 2nd edition (ICSD-2; American Academy of Sleep Medicine, 2005).

A problem arises when measuring depressed mood as an outcome variable in a context where sleep, sleep schedules, or circadian rhythms are independent variables and sleep items are included in the mood measure. Similarly, disturbed sleep symptoms influence reports of depressed mood within clinical settings. This challenge is not unique to sleep. The frequent co-occurrence of vegetative symptoms with depressed mood, for example, creates problems for those assessing depression in the context of medical disorders (Volk, Pace, & Parchman, 1993). As with sleep, standard depressed mood scales often contain items that address such vegetative symptoms. A solution was to propose and validate the Beck Depression Inventory for Primary Care (Beck, Guth, Steer, & Ball, 1997), a depressed mood measure void of confounding items assessing vegetative symptoms.

The current analysis examines whether a brief depressed mood scale validated for adolescents (Kandel & Davies, 1982) retains reliability when items addressing sleep symptoms are removed. We also evaluate the item content of the scale to assess the scale's convergent validity without the sleep-related items and the impact of removing items on the association between the depressed mood scale and the Center for Epidemiologic Studies Depression Scale (CES-D).

METHODS

Participants

In September 2010, 182 first-year Brown University students completed daily online sleep diaries and six biweekly surveys; 178 students met inclusion criteria for the current analyses by completing at least two biweekly online surveys. Ages were 18-21 years (mean = 18.1 years; $SD = 0.4$) with 108 females. Ethnic and racial make-up was 89 Caucasians, 35 Asians, 14 African American, 18 multiracial, 16 Hispanic, 3 other, and 3 declined to report.

Procedures

Students logged into a secure online portal each evening to complete their sleep diaries. Biweekly surveys querying mood, workload (e.g., exams, homework), and medication use during

the two weeks prior were available from Friday afternoon to Sunday evening. Students selected which night to complete the survey and completed it one time at weeks 2, 6, 8, and 10. In addition to the biweekly assessment, the survey was presented on two consecutive days at weeks 4 and 12. The CES-D was administered during week 9. The software IllumeTM (DatStat, Inc.) was used to create the surveys and collect the data online. Analyses were performed using SPSS 18 (IBM®). One hundred twenty-eight students completed all six surveys with an average biweekly survey completion rate of 5.51 out of six opportunities ($SD = 0.94$); for the consecutive day test-retest surveys, 145 students participated in at least one opportunity and 78 participated in both opportunities with an average completion rate of 1.54 out of two opportunities ($SD = 0.5$).

Measures

The Kandel and Davies (1982) depressive mood scale included on the biweekly surveys is a 6-item self-report measure used to assess depressed mood. Individual items are listed in Table 1; response options for each item were *Not at all*, *Somewhat*, or *Much*. The original scale showed good reliability and validity to assess depressed mood of adolescents aged 14-18 years.

The Center for Epidemiologic Studies — Depressed Mood Scale (CES-D; Radloff, 1977) is a 20-item scale. CES-D items query various depressed mood symptoms experienced during the past week including poor appetite, feeling depressed, hopelessness, and “restless” sleep. Responses on each item include *Rarely or none of the time* (less than one day), *Some or a little of the time* (one-two days), *Occasionally or a moderate amount of the time* (three-four days), or *Most or all of the time* (five-seven days). Total scores on the CES-D range from 20-80 with higher scores indicating more depressive symptomatology. Our sample showed adequate internal consistency (Cronbach’s alpha = .91).

Data Analyses

For these analyses, the Kandel and Davies scale was examined as 1) the full 6-item scale, 2) a 5-item scale that excluded the sleep item (“trouble going to sleep or staying asleep”), and 3) a 4-item scale that excluded the sleep and tired (“feeling too tired to do things”) items. While “trouble going to sleep or staying asleep” overtly describes disturbed sleep, “feeling too tired” could be considered unrelated to sleep; however, the average lay-person may report “feeling tired” as a state of “sleepiness.”

Intraclass correlation (ICC) analyses evaluated the test-retest reliability of the three versions of the Kandel and Davies (1982) scale cumulatively across the six biweekly administrations. ICC values were computed for weeks 2 and 4; weeks 2, 4, and 6; weeks 2, 4, 6, and 8; and so forth. ICC values were also computed for the two consecutive-day administrations (weeks 4 and 12) to assess short-term test-retest reliability. Corrected item-total correlation and Cronbach’s alpha were examined for the 6-item scale, 5-item scale, and 4-item scale at all six biweekly administrations. In addition, inter-measure correlations with the CES-D were examined for the 6-item scale, 5-item scale, and 4-item scale at each administration.

TABLE 1
Psychometric characteristics of the three scales

Scale items (6-item scale)	Corrected item-total correlations					
	Week 2	Week 4	Week 6	Week 8	Week 10	Week 12
1. Feeling too tired to do things	.28	.46	.38	.44	.43	.41
2. Having trouble going to sleep or staying asleep	.35	.27	.25	.20	.30	.38
3. Feeling unhappy, sad, or depressed	.58	.64	.63	.59	.64	.61
4. Feeling hopeless about the future	.56	.55	.58	.63	.66	.57
5. Feeling nervous or tense	.57	.63	.57	.59	.73	.45
6. Worrying too much about things	.62	.63	.64	.62	.62	.65
Scale items (5-item scale)	Week 2	Week 4	Week 6	Week 8	Week 10	Week 12
1. Feeling too tired to do things	.26	.45	.36	.41	.43	.35
2. Feeling unhappy, sad, or depressed	.55	.66	.65	.61	.65	.60
3. Feeling hopeless about the future	.56	.56	.62	.63	.66	.56
4. Feeling nervous or tense	.59	.63	.59	.64	.72	.48
5. Worrying too much about things	.65	.66	.63	.63	.64	.69
Scale items (4-item scale)	Week 2	Week 4	Week 6	Week 8	Week 10	Week 12
1. Feeling unhappy, sad, or depressed	.56	.66	.67	.62	.67	.60
2. Feeling hopeless about the future	.60	.57	.59	.64	.63	.59
3. Feeling nervous or tense	.58	.63	.60	.63	.71	.48
4. Worrying too much about things	.66	.65	.64	.64	.66	.70
Scale	Cronbach's alpha					
	Week 2	Week 4	Week 6	Week 8	Week 10	Week 12
6-item scale	.75	.78	.76	.76	.80	.76
5-item scale	.75	.80	.79	.80	.82	.76
4-item scale	.79	.81	.81	.81	.84	.78

Note. Bold values reflect an ICC value of < .50.

RESULTS

Biweekly Reliability Analyses

The ICC values across the study weeks (e.g., weeks 2 to 4, weeks 2 to 6, weeks 2 to 8) are illustrated in Figure 1, showing reliability estimates ranging from .65 to .68 for the full 6-item scale, .63 to .68 for the 5-item scale, and .63 to .67 for the 4-item scale. Thus, the ICC values for repeated assessments were moderately similar across time regardless of item exclusion. ICC assessment of stability across all six biweekly assessments showed reliability estimates of .65 for the full 6-item scale, .63 for the 5-item scale, and .63 for the 4-item scale.

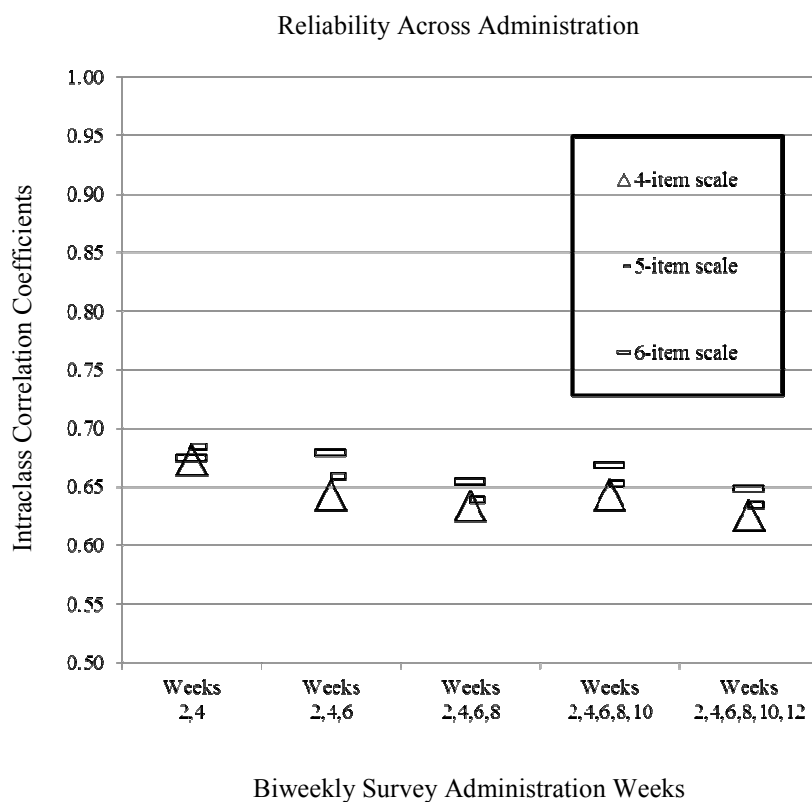


FIGURE 1
Reliability across administrations.
This figure illustrates the intraclass correlation coefficients for a) the full 6-item scale, b) a 5-item scale, and c) a 4-item scale consecutively as the study progressed from weeks 2 to 12.

Consecutive-Day Reliability Analyses

The ICC reliability estimates for consecutive-day assessments were .91 (6-item), .92 (5-item), and .91 (4-item) at week 4 and .88 (6-item), .87 (5-item), and .88 (4-item) at week 12. Similar to the biweekly reliability analyses, the ICC values were stable for the consecutive-day assessments regardless of assessment week.

Item Analyses

The corrected item-total correlations for each item in the three scales (6-item, 5-item, and 4-item scales) at every biweekly assessment are shown in Table 1. For the 6-item scale, all correlations were greater than .55 for every assessment, except the sleep (item 2; $r = .20$ to $.38$) and tired (item 1; $r = .28$ to $.46$) items at all assessments and the “feeling nervous or tense” item (item 5; $r = .45$) at week 12.

Corrected item-total correlations for the 5-item scale yielded correlations similar to the 6-item scale. All correlations were .55 or greater for every assessment, except the tired item ($r = .26$ to $.45$) at all assessments and the nervous item ($r = .48$) at week 12.

Corrected item-total correlations for the 4-item scale yielded correlations of .56 or greater, with the exception of the nervous item on the last administration (week 12). Cronbach’s alpha was equal to or higher than .75 for all three scales at all six administrations. For the 4-item scale specifically, Cronbach’s alpha was .78 to .84 for each administration.²

Inter-Measure Analyses

Correlations for the CES-D with each assessment for the 6-item scale, 5-item scale, and 4-item scale are shown in Table 2. Correlations ranged from $r = .57$ to $.76$ for the 6-item scale, from $r = .58$ to $.77$ for the 5-item scale, and from $r = .55$ to $.77$ for the 4-item scale. Higher correlations were observed closer in proximity to the CES-D administration, which occurred during week 9. All three scales showed comparable association with the CES-D total scores for each biweekly assessment.

TABLE 2
 Comparison of depressed mood scales

Scale	Correlation with CES-D Total Score (Week 9)					
	Week 2	Week 4	Week 6	Week 8	Week 10	Week 12
6-item scale	.58	.61	.70	.73	.76	.57
5-item scale	.59	.63	.71	.74	.77	.58
4-item scale	.55	.62	.68	.75	.77	.59

Note. $p < .001$. Bold values indicate the correlations for the biweekly surveys completed closest in proximity to the CES-D administration.

DISCUSSION

The lack of change in ICC values, the association between the CES-D and the depressive mood scale when sleep items are removed support removing the sleep items from the Kandel and Davies’ (1982) depressed mood scale when sleep is an outcome variable of interest. This conclusion is similar to that used within medical populations, where reliable measures of depression ha-

ve been adapted to exclude confounding vegetative symptoms (Beck et al., 1997; Volk et al., 1993).

Stability in participants' responses across consecutive-day and biweekly administrations was comparable for the full depressed mood scale and when sleep-related items were removed. The nervous item was retained because its content is similar to that of the other items and correlations on the other five administrations were consonant with the other mood items on the 6-item scale ($r = .57-.73$). Correlation analyses with the CES-D showed all three scales were associated with the CES-D total score, particularly the assessments at weeks 8 and 10 that bracketed the CES-D administration. Our findings indicate that using the 4-item scale is a reliable measure that is consistently correlated with the CES-D. The resulting depressed mood score is free of potential confounds that sleep questions introduce.

A limitation of these data was a violation to the normal distribution requirement for items 1-5 of the full 6-item scale in the current analyses (Table 1). Item 1 showed a leptokurtic distribution and items 2 through 5 were positively skewed. This violation was offset by the large sample size with the methods used, which are robust against violations of normality assumptions. Another concern was the high frequency of zeros reported on the sleep item, which likely affected its total item correlation. Although this high frequency of zeros was also reported on the hopelessness item, however, this item maintained a similar correlation to the other items on the mood scale at every administration. Thus, the low total item correlation value is likely driven by the sleep item being distinct from the remaining scale items and not from a high percentage of zeroes (zero was equal to a response of *not at all*), which was a common response regardless of other responses on the survey. The current study used the CES-D (Radloff, 1977) to assess depressive mood instead of the recently revised CESD-R (Eaton, Mutaner, Smith, Tien, & Ybarra, 2004; Van Dam & Earleywine, 2011). The original CES-D exhibits good psychometric properties. The revised version was more recently developed with a focus on DSM-IV-specific symptomology. Fewer reliability and validity studies have been done using the CESD-R, and debate is ongoing about the benefit, or potential consequences, of removing positive items (Van Dam & Earleywine, 2011; Wood, Taylor, & Joseph, 2010).

The findings confirm the reliability of the Kandel and Davies (1982) scale when the sleep-related items are excluded. In other words, removal of sleep items does not affect the reliable assessment of depressed mood with this measure, and this procedure limits potential confounds when investigators wish to untangle sleep, sleep schedules, or circadian rhythms from depressed mood. Clinicians also face the challenge of reliably and quickly assessing depressed mood. A quick assessment is particularly important in college populations due to the burden on the small number of health professionals who most oversee the psychological and medical health of many students. Teasing apart symptoms of depressed mood versus sleep disturbance requires time, and a few shorter assessments with minimal overlap would reduce clinician burden by allowing clinicians to more quickly identify the symptoms troubling the patient while excluding concerns that are not present or paramount.

While the scope of this manuscript does not extend to the reliability of insomnia scales, several scales are available for clinical or research use. Examples of these measures include the Insomnia Severity Index (ISI; Bastien, Vallières, & Morin, 2001), Women's Health Initiative Insomnia Rating Scale (WHIIRS; Levine et al., 2003), Insomnia Symptom Questionnaire (ISQ; Okun et al., 2009), and Brief Insomnia Questionnaire (BIQ; Kessler et al., 2010).

NOTES

1. Support received from the National Institute of Mental Health (MH079179) and the Sleep Research Society Foundation Elliot D. Weitzman, M.D. Research Grant. Tifenn Raffray received support from the Institut Servier, France. The authors have no other financial disclosures or conflicts of interests. We are indebted to our research staff, including David Baker, David Bushnell, Erin Campopiano, Gretchen Su-
roff, James Bass, Elyn Ferriter, Katie Esterline, Michelle Loxley, and Caroline Gredvig-Ardito.
2. A Principal Axis Factor (PAF) analysis with a Direct Oblimin (oblique) rotation of the 6-item scale was conducted on the first assessment of data. The six items revealed that all items correlated at least .3 with at least one other item. In addition, the Kaiser-Meyer-Olkin measure of sampling adequacy was above the recommended value of .6 (KMO = .77, $p < .001$), and Bartlett's test of sphericity was significant ($\chi^2 = 231.56, p < .001$). Inclusion of each item in the factor analysis was also supported by all the diagonals of the anti-image correlation matrix being over .5. When loadings less than .30 were excluded, the analysis yielded a one-factor solution with a simple structure (factor loadings = .67 to .75; communalities = .47 to .57). The sleep and tired items were dropped from the factor solution supporting the removal of these items from the brief scale.

REFERENCES

- American Academy of Sleep Medicine (2005). *International classification of sleep disorders: Diagnostic and coding manual* (2nd ed.). Westchester, IL: American Academy of Sleep Medicine.
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.). Washington, DC: Author.
- Bastien, C. H., Vallières, A., & Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine, 2*(4), 297-307. doi:10.1016/S1389-9457(00)00065-4
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- Beck, A. T., Guth, D., Steer, R. A., & Ball, R. (1997). Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behaviour Research and Therapy, 35*, 785-791. doi:10.1016/S0005-7967(97)00025-9
- Eaton, W. W., Mutaner, C., Smith, C., Tien, A., & Ybarra, M. (2004). Center for Epidemiologic Studies Depression Scale: Review and revision (CESD and CESD-R). In M. E. Maruish (Ed.), *Use of psychological testing for treatment planning and outcomes assessments* (3rd ed., pp. 363-377). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*, 56-62.
- Kandel, D. B., & Davies, M. (1982). Epidemiology of depressive mood in adolescents: An empirical study. *Archives of General Psychiatry, 39*, 1205-1212.
- Kessler, R. C., Coulouvrat, C., Hajak, G., Lakoma, M. D., Roth, T., Sampson, N., Shahly, V., et al. (2010). Reliability and validity of the brief insomnia questionnaire in the America insomnia survey. *Sleep, 33*(11), 1539-1549.
- Levine, D. W., Kripke, D. F., Kaplan, R. M., Lewis, M. A., Naughton, M. J., Bowen, D. J., & Shumaker, S. A. (2003). Reliability and validity of the Women's Health Initiative Insomnia Rating Scale. *Psychological Assessment, 15*(2), 137-148. doi:10.1037/1040-3590.15.2.137
- Manber, R., & Chambers, A. S. (2009). Insomnia and depression: A multifaceted interplay. *Current Psychiatry Reports, 11*, 437-442. doi:10.1007/s11920-009-0066-1
- Okun, M. L., Kravitz, H. M., Sowers, M. F., Moul, D. E., Buysse, D. J., & Hall, M. (2009). Psychometric evaluation of the Insomnia Symptom Questionnaire: A self-report measure to identify chronic insomnia. *Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine, 5*(1), 41-51.
- Pigeon, W. R., & Perlis, M. L. (2007). Insomnia and depression: Birds of a feather? *International Journal of Sleep Medicine, 1*, 82-91.
- Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401. doi:10.1177/014662167700100306
- Taylor, D. J. (2008). Insomnia and depression. *Sleep, 31*, 447-448.
- Van Dam, N. T., & Earleywine, M. (2011). Validation of the Center for Epidemiologic Studies Depression Scale-Revised (CESD-R): Pragmatic depression assessment in the general population. *Psychiatry Research, 186*(1), 128-132. doi:10.1016/j.psychres.2010.08.018
- Volk, R. J., Pace, T. M., & Parchman, M. L. (1993). Screening for depression in primary care patients: Dimensionality of the Short Form of the Beck Depression Inventory. *Psychological Assessment, 5*, 173-181.

Wood, A. M., Taylor, P. J., & Joseph, S. (2010). Does the CES-D measure a continuum from depression to happiness? Comparing substantive and artifactual models. *Psychiatry Research*, *177*(1-2), 120-123. doi:10.1016/j.psychres.2010.02.003

TPM[®]
