

Evaluation of Two Circadian Rhythm Questionnaires for Screening for the Delayed Sleep Phase Disorder

Min Kyu Rhee¹, Heon-Jeong Lee²✉, Katharine M. Rex³ and Daniel F. Kripke³✉

¹Department of Psychology, Gyeongsang National University, Jinju, Republic of Korea

²Department of Psychiatry, Korea University College of Medicine, Seoul, Republic of Korea

³Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA

Objective Delayed sleep phase disorder (DSPD) is a condition in which patients often fall asleep some hours after midnight and have difficulty waking up in the morning. Circadian chronotype questionnaires such as Horne-Östberg Morningness-Eveningness Questionnaire (MEQ) and Basic Language Morningness (BALM) scale have been used for screening for DSPD. This study was to evaluate these two chronotype questionnaires for screening of DSPD.

Methods The study samples were 444 DSPD and 438 controls. Cronbach's alpha coefficient was calculated to evaluate for internal consistency. An exploratory factor analysis was conducted using principal-axis factoring. The diagnostic performance of a test was evaluated using Receiver Operating Characteristic (ROC) curve analysis. A discriminant function analysis was also performed.

Results For internal consistency, Cronbach's alpha of 0.898 for BALM was higher than the 0.837 for MEQ, though both have acceptable internal consistency. BALM has better construct validity than the MEQ because some MEQ items measure different dimensions. However, when we evaluated the efficiency of two questionnaires for DSPD diagnosis by using the ROC curve, the BALM was similar to the MEQ. In a discriminant analysis with the BALM to classify the two groups (DSPD vs. normal), 6 items were identified that resulted in good classification accuracy. Upon examination of the classification procedure, 94.2% of the originally grouped cases were classified correctly.

Conclusion These findings suggest that the BALM has better psychometric properties than the MEQ in screening and discriminating DSPS. **Psychiatry Investig 2012;9:236-244**

Key Words Delayed sleep phase disorder, Circadian rhythm questionnaire, Validity, Chronotype.

INTRODUCTION

Chronotype is an individual difference reflecting the time of day at which individuals are "at their best".^{1,2} Chronotype is often assessed by means of self-rating questionnaires. The first and the most widely used chronotype questionnaire is the Morningness-Eveningness Questionnaire (MEQ).³ Some of

the criticisms addressed to the MEQ are that the total score may not be appropriate to measure a multidimensional construct,⁴ and that small subsets of items may convey most of the total variance of the measure.^{5,6} Similar criticisms have been raised for the Circadian Type Questionnaire (CTQ) developed by Greenwood.⁷ Smith et al.⁸ developed a new questionnaire called the Composite Scale of Morningness (CSM) that is composed by the best items of the MEQ (nine items) and the CTQ (four items). Because CSM was criticized for language difficulty due to British English, it was simplified into a 'basic language morningness' (BALM) scale at a seventh grade (12-13 years) reading level.⁹ Both MEQ and BALM are widely used in evaluating morningness in US. The original Horne-Östberg scale has also been rephrased into a simplified American idiom,¹⁰ which was used for this study.

Delayed sleep phase disorder (DSPD) is a circadian rhythm sleep disorder. Typical DSPD patients tend to fall asleep some hours after midnight and have difficulty waking up in the morn-

Received: May 22, 2012 Revised: June 7, 2012

Accepted: June 7, 2012 Available online: September 6, 2012

✉ Correspondence: Daniel F. Kripke, MD
Department of Psychiatry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0605, USA
Tel: +1-858-222-2934, E-mail: dkripke@ucsd.edu

✉ Correspondence: Heon-Jeong Lee, MD, PhD
Department of Psychiatry, Korea University Anam Hospital, Korea University College of Medicine, 126-1 Anam-dong 5-ga, Seongbuk-gu, Seoul 136-705, Republic of Korea
Tel: +82-2-920-6721, Fax: +82-2-929-7679, E-mail: leehjeong@korea.ac.kr

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ing. A prevalence of DSPD in the general adult population, equally distributed among women and men, has been reported at approximately 0.15%,^{11,12} but Ando et al.¹³ found that the prevalence of mild DSPD symptoms might be much higher. Patients with DSPD may be unable to fall asleep early enough to rest adequately before it is time for school or work, or they may be frequently unable to arrive at school or work on time. Consequently, DSPD can be a disabling and socially isolating condition, unless the patients are able to fit their habits into an accommodating social milieu.

Although sleep logs and actigraphy monitoring are helpful, there is no diagnostic test for DSPD, and diagnosis is based on the sleep history. Chronotype questionnaires such as MEQ and BALM have been used for screening for DSPD as a way of obtaining the history. However, there has been no information which one is better for the assessment for DSPD. We have performed genetic studies of DSPD and collected large samples of DNA from DSPD volunteers and matched controls. The purpose of this paper was evaluation of the two chronotype questionnaire for screening of DSPD.

To assess the psychometric characteristics of BALM and MEQ, reliability, validity and diagnostic utility were examined for each questionnaire. Concretely, we compared the two scales on the item-total correlation, scale coefficient alpha (internal consistencies), construct validity (factor analyses), criterion-related validity and classification accuracy (ROC curve analysis). Also discriminant analysis was used to determine which items were the best predictors of DSPD.

METHODS

Participant recruitment

The study participants were recruited for a case-control genetic study of DSPD. The recruitment process was described in previous reports.^{14,15} A brief description is as follows. In the initial part of the study, recruitment was limited to the Southern California region and was later expanded to other states of the US. Recruitment of the sample took place between June, 2004 and May, 2010. Some data from the sample was published previously.^{14,15} Recruitment of DSPD participants utilized contacts with sleep physicians, media contacts, UCSD minority outreach programs, and internet advertising. Only participants 25 years of age or older were accepted (with a few exceptions). DSPD participants completed the questionnaires and wore a wrist actigraph for 2 weeks. Normal control volunteers were recruited by word-of-mouth, community meetings, the internet, and by a campus poster. Control recruitment was targeted, so far as possible, to match the ancestry of the case series. Control participants completed the questionnaires and contributed a sample of blood or saliva, but they were not

asked to wear the actigraphs or to provide sleep logs. The principal investigator (DFK) reviewed the record of each participant volunteering as a control. Based on their questionnaires, control volunteers were retrospectively rated as 1) certain DSPD, 2) possible DSPD, 3) neither, 4) possible ASPS, or 5) certain ASPS. Only those with ratings 3 or 4 were included among the normal controls in this analysis. All the subjects received an explanation of the study and signed written informed consent.

DSPD diagnosis

Once all data were assembled, the principal investigator (DFK) reviewed the record of each participant who had volunteered as a DSPD case and recorded the participant's DSPD classification as 1) absolutely certain, 2) fairly certain, 3) questionable, 4) unlikely, or 5) very doubtful. The initial criterion for classification was the MEQ, recognizing that the criterion for definite evening type of <30 was too strict for the San Diego population.¹⁴ Confirmatory classification criteria included the score on the BALM, reported prior-week and adult-life bedtimes and awakening times, the actigraphic recordings, and whether the participant reported going to sleep "somewhat later" or "much later" than most people their age, both as a child and as an adult. Whether the participant reported distress about falling asleep, reported related social or vocational problems, or had sought medical attention for a sleep problem was also considered. The consistency of the data supporting a classification of DSPD was evaluated, together with the presence of depression, other mental illnesses, or other sleep disorders which might confuse the classification. However, if depression or other disorders had their first incidence after the onset of a pattern of delayed sleep and did not appear to be causing the delay, these disorders were not considered exclusionary. As many DSPD patients cannot consistently report to work by 8-9 AM, evening or night shift work was not considered exclusionary if the history indicated that the delay in sleep occurred before shift work was adopted, and the delay tended to persist when the participant was off work.

Statistical analysis

Internal consistency

To determine internal consistency, Cronbach's alpha coefficients were calculated for the BALM and MEQ. Cronbach's alpha coefficient was calculated to determine the degree of homogeneity among the items in scale. A Cronbach's alpha of 0.80 may be considered acceptable. Between 0.80 and 0.89, the level of item consistency is good; and when it is 0.90 and above, it is excellent.¹⁶

Comparison of validity

Two types of validity were examined: construct and criterion validity. First, the internal structure or construct validity of each subscale within the BALM and MEQ was evaluated based on an exploratory factor analysis (EFA). Second, criterion validity was examined by computing correlations between the two scales and DSPD diagnoses.

For construct validity, an EFA was conducted using principal-axis factoring in SPSS (Version 15) on each scale to explore underlying dimensions. To minimize subjectivity, it was decided to apply the following decision criteria: 1) Kaiser's eigenvalue rule which is a psychometric criterion, 2) Cattell's scree plot and 3) the interpretability criterion.¹⁷ In EFA, factor loadings are generally considered to be meaningful when they exceed 0.30 or 0.40.¹⁷ To determine inclusion of an item in a factor dimension, a score above 0.35 on a primary loading of items after rotation was used as the cut off. In order to show criterion-related validity, each of two circadian rhythm scale scores (interval variable) was compared to a dichotomous variable, DSPD diagnosis using an eta (η) correlation coefficient.

ROC curve analysis

The diagnostic performance of a test, or its accuracy in discriminating diseased cases from normal cases is evaluated using Receiver Operating Characteristic (ROC) curve analysis. Receiver-operating characteristic (ROC) plots provide a pure index of accuracy by demonstrating the limits of a test's ability to discriminate between alternative states of health over the complete spectrum of operating conditions,¹⁸ over a range of threshold levels.

With ROC, the effectiveness of an instrument is assessed by evaluating the accuracy of discrimination between two groups. An area under the curve (AUC) of 1 defines a perfect test, whereas an area of 0.5 represents a completely inefficient measure; ROC areas of 0.80-0.90 are considered good discriminators, and curves of 0.90-1 are considered excellent.¹⁹ We have tried to evaluate whether BALM and MEQ performed better for evaluating and making diagnoses of DSPD, with the caution that the MEQ was generally more influential a priori in influencing the diagnostic rater, because MEQ ratings were more familiar. The efficiency of the two assessment tools for DSPD diagnosis was estimated by using ROC Curve non-parametric analysis. This curve, used to assess the accuracy of diagnosis criteria, offers a graph of true positive rate, sensitivity vs. false positive rate, in other words, 1.0-specificity.

Discriminant function analysis

A discriminant function analysis (DFA) was performed in order to determine whether DSPD case could be reliably classified from a set of predictors (circadian rhythm scale items)

and whether any of the scale items demonstrated differential item functioning as a function of group. That is, DFA enabled us to determine the relative importance of the dependent variables (DVs) in discriminating among DSPD groups. In this study, the DVs (circadian rhythm scale items) were treated as predictors in order to examine how they were able to predict group membership for the two DSPD groups (DSPD and normal). Univariate F tests were then calculated in order to determine the importance of each independent variable (IV) in forming the discriminant functions.

Examining the Wilk's Lambda values for each of the predictors revealed how important the IV was to the discriminant function, with smaller values representing greater importance.

Statistical packages

All of the analyses were performed using SPSS for Windows version 15 (SPSS, Chicago, IL, USA), with the cutoff for statistical significance set at $p < 0.05$.

RESULTS

Subjects

By the recruitment process described above, we recruited a case series of 387 DSPD along with 361 controls. Their mean age was 38.41 (SD=12.42). The gender distribution was 67.1% female and 32.9% male.

Internal consistency

Table 1 and 2 show item-total statistics and Cronbach's alpha for the two scales, respectively. Internal consistency alphas were 0.898 for BALM and 0.837 for MEQ. In terms of standards, Nunnally recommends that coefficient alpha be 0.70 or higher in basic research and 0.90 or higher in applied settings where clinical decisions are based on test scores.²⁰ Both scales had acceptable internal consistency. However, although Cronbach's alpha is usually higher for the large items scale than the small items scale because the number of items affects the magnitude of the coefficient,²¹ the alpha coefficient of the BALM was higher than that of the MEQ in spite of small item numbers of BALM. The corrected item-total correlations (homogeneity index) were greater than 0.40 (considered to be a marker of good coefficients) in all items of the BALM scale, but for the MEQ, seven of items (nos. 3, 6, 8, 12, 13, 14, 16) were lower than 0.40 (Table 1 and 2).

Comparison of validity

Construct validity

For the EFA, orthogonally rotated factor solutions were used. Table 3 and 4 give the factor loading matrix for two

Table 1. BALM scale item-total statistics and Cronbach's alpha

Item	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Squared multiple correlation	Cronbach's alpha if item deleted
1. Thinking of your "feeling best" times of day, what time would you get up if you were completely free to plan your day?	35.5594	43.332	0.631	0.487	0.889
2. Thinking only of your own "feeling best" times of day, what time would you go to bed if you were completely free to plan your evening?	35.7306	47.808	0.433	0.421	0.897
3. Under free time conditions, how easy do you find getting up in the morning?	35.9521	45.158	0.600	0.465	0.890
4. If you go to sleep at night at a regular time, how wide awake do you feel during the first half hour after getting up in the morning?	36.1826	45.083	0.589	0.572	0.891
5. If you go to sleep at night at a regular time, during the first half hour after getting up in the morning, how tired or rested do you feel?	36.1256	46.586	0.518	0.489	0.894
6. In your free time you have decided to do some physical exercise A friend says do this one hour twice a week and the best time is 7:00-8:00 in the morning. Thinking of nothing else but your own "feeling best" times, how do you think you would do at that time?	35.9658	44.276	0.632	0.433	0.889
7. Suppose you have nothing but free time, at what time in the evening do you feel tired and need to get some sleep?	35.7032	47.463	0.467	0.441	0.896
8. You want to be at your top performance for a test [for a job promotion, for example] which you know is going to take two hours and will completely tire you out mentally. You are completely free to plan your day. Thinking only of your own "feeling best" times, which ONE of the four testing times would you choose?	35.5068	46.548	0.560	0.362	0.892
9. "Morning people" wake up early and are wide awake fast, but then feel tired in the evenings and like to go to bed before it gets late. "Evening people" take some time before they get going in the mornings, but like to stay up late at night. Which ONE of these types do you think you are?	36.1027	42.628	0.737	0.693	0.883
10. When would you like to get up (if you have a full day's work-8 hours-ahead of you, for example) if you were completely free to set your work time?	36.4269	44.112	0.594	0.472	0.891
11. Suppose that you did not work, but you still always had to get up at 6:00 in the morning. What do you think it would be like for you?	36.1393	43.132	0.677	0.492	0.886
12. If you could go to sleep at a regular time every night, how much time would it take to "clear your head" in the morning after getting up from a night's sleep?	35.6758	45.172	0.571	0.420	0.892
13. If you were free to set your time, please mark how much you would be a morning or evening active person?	36.0251	42.331	0.758	0.706	0.882
Reliability Statistics Cronbach's Alpha	0.898				
BALM: basic language morningness					

Table 2. Item-total statistics in morningness-eveningness questionnaire

Items	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Squared multiple correlation	Cronbach's alpha if item deleted
1. Approximately what would you get up if you were entirely free to plan your day?	53.4462	80.078	0.645	0.527	0.820
2. Approximately what time would you go to bed if you were entirely free to plan your evening?	53.7941	83.558	0.545	0.445	0.827
3. If you usually have to get up at a specific time in the morning, how much do you depend on an alarm clock?	54.6979	82.555	0.364	0.257	0.832
4. How easy do you find it to get up in the morning (when you are not awakened unexpectedly)?	53.9817	82.156	0.580	0.505	0.824
5. How alert do you feel during the first half hour after you wake up in the morning?	53.9817	83.463	0.476	0.436	0.828
6. How hungry do you feel during the first half hour after you wake up?	55.0183	89.234	0.037	0.058	0.846
7. During the first half hour after you wake up in the morning, how do you feel?	54.1991	83.637	0.472	0.455	0.828
8. If you had no commitments the next day, what time would you go to bed compared to your usual bedtime?	54.3707	84.041	0.378	0.248	0.831
9. You have decided to do physical exercise. A friend suggests that you do this for one hour twice a week, and the best time for him is between 7-8 A.M. Bearing in mind nothing but your own internal "clock," how do you think you would perform?	53.8352	81.129	0.532	0.372	0.825
10. At approximately what time in the evening do you feel tired, and as a result, in need of sleep?	53.4485	83.996	0.442	0.374	0.829
11. You want to be at your peak performance for a test that you know is going to be mentally exhausting and will last two hours. You are entirely free to plan your day. Considering only your "internal clock," which one of the four testing times would you choose?	52.1419	74.131	0.585	0.438	0.820
12. If you got into bed at 11 P.M., how tired would you be?	54.2426	84.170	0.214	0.179	0.843
13. For some reason you have gone to bed several hours later than usual, but there is no need to get up at any particular time the next morning. Which one of the following are you likely to do?	54.3982	83.034	0.346	0.257	0.833
14. One night you have to remain awake between 4-6 A.M. in order to carry out a night watch. You have no commitments the next day. Which one of the alternatives would suit you best?	54.0160	84.924	0.293	0.133	0.835
15. You have two hours of hard physical work. You are entirely free to plan your day. Considering only your internal "clock," which of the following times would you choose	53.5286	83.773	0.457	0.344	0.829
16. You have decided to do physical exercise. A friend suggested that you do this for one hour twice a week. The best time for her is between 10-11 P.M. Bearing in mind only your internal "clock," how well do you think you would perform?	54.5538	85.578	0.198	0.148	0.841
17. Suppose you can choose your own work hours. Assume that you work a five-hour day (including breaks), your job is interesting, and you are paid based on your performance. At approximately what time would you choose to begin?	53.3341	82.838	0.525	0.348	0.826
18. At approximately what time of day do you usually feel your best?	53.5606	81.875	0.627	0.487	0.823
19. One hears about "morning types" and "evening types." Which one of these types do you consider yourself to be?	53.1854	66.303	0.700	0.546	0.813
Reliability Statistics	Cronbach's Alpha	0.837			

Table 3. Matrix of orthogonally rotated factor loading for BALM scale

Item	Factor			Communalities
	1	2	3	
BALM10	0.700	0.108	0.174	0.532
BALM1	0.639	0.175	0.270	0.512
BALM11	0.635	0.312	0.188	0.535
BALM13	0.629	0.392	0.310	0.645
BALM9	0.620	0.385	0.281	0.611
BALM6	0.604	0.311	0.121	0.476
BALM8	0.571	0.231	0.110	0.392
BALM4	0.173	0.841	0.128	0.753
BALM5	0.192	0.721	0.029	0.558
BALM3	0.346	0.605	0.103	0.496
BALM12	0.324	0.573	0.110	0.445
BALM2	0.194	0.093	0.802	0.689
BALM7	0.283	0.097	0.679	0.550
Eigenvalues	3.180	2.530	1.484	7.194
% of variance explained	24.46	19.46	11.42	55.34

BALM: basic language morningness

scales. As shown in Table 3, three factors with eigenvalue of 1.0 or above were extracted from the BALM by using principal axis factoring with varimax rotation. A 3-factor solution accounted for 55.34% of the variance: Factor I (eigenvalue, 3.18; variance, 24.46%); Factor II (eigenvalue, 2.53; variance, 19.46%), and Factor III (eigenvalue, 1.484; variance, 11.42%). The first factor was defined by 7 items (nos. 1, 6, 8, 9, 10, 11, 13) with factor loading above 0.50. The items of this factor was congruent with factor 1 (morningness/effort) in the Brown study.⁹ The second factor was defined by 4 items (nos. 3, 4, 5, 12) with loading above 0.50, in common with factor 2 (morning alertness) in the Brown study.⁹

The third factor (evening) had 2 items with loading above 0.60 that refer to activities and affect in the evening as defined by the Brown study.⁹ Regarding the alpha values for the subscales of BALM, factor 1 (alpha=0.880), 2 (alpha=0.826), and 3 (alpha=0.759) subscales were above Nunnally's critical value of alpha >0.70.²⁰

From the MEQ, five factors with eigenvalue of 1.0 or above were extracted by using principal axis factoring and orthogonally rotated by using varimax rotation (Table 4). Of the 19 items, 15 items loaded above 0.40 on one of the five extracted

Table 4. Matrix of orthogonally rotated factor loading for Morningness-Eveningness Questionnaire

Item	Factor					Communality
	1	2	3	4	5	
MEQ11	0.670	0.166	0.203	0.081	0.003	0.524
MEQ15	0.631	0.041	0.015	0.137	-0.026	0.419
MEQ18	0.615	0.228	0.217	0.154	0.090	0.509
MEQ17	0.576	0.139	0.117	0.143	0.103	0.396
MEQ19	0.574	0.322	0.299	0.217	0.113	0.582
MEQ09	0.517	0.261	0.084	0.208	0.008	0.386
MEQ14	0.295	0.123	0.122	-0.046	0.175	0.150
MEQ07	0.167	0.757	0.056	0.141	-0.027	0.625
MEQ05	0.301	0.666	0.029	0.059	-0.009	0.539
MEQ04	0.316	0.626	0.110	0.302	-0.113	0.607
MEQ08	0.123	0.286	0.172	0.234	0.272	0.255
MEQ10	0.208	0.048	0.683	0.122	0.036	0.529
MEQ02	0.251	0.202	0.553	0.168	0.345	0.557
MEQ12	0.073	-0.005	0.504	0.019	-0.135	0.278
MEQ16	0.061	0.044	0.374	-0.021	0.006	0.146
MEQ13	0.208	0.067	0.003	0.609	0.095	0.427
MEQ03	0.119	0.194	0.071	0.551	0.027	0.361
MEQ01	0.448	0.253	0.233	0.369	0.496	0.700
MEQ06	-0.003	0.097	0.075	-0.018	-0.226	0.066
Eigenvalues	2.847	1.938	1.491	1.185	0.595	8.056
% of variance explained	14.98	10.20	7.85	6.24	3.13	42.40

MEQ: Morningness-Eveningness Questionnaire

factors. A 5-factor solution accounted for 42.40% of the variance: Factor I (eigenvalue, 2.847; variance, 14.98%); Factor II (eigenvalue, 1.938; variance, 10.20%), Factor III (eigenvalue, 1.491; variance, 7.85%), Factor IV (eigenvalue, 1.185; variance, 6.24%), and Factor V (eigenvalue, 0.595; variance, 3.13%).

The first factor was composed of 6 items (nos. 9, 11, 15, 17, 18, 19). Four of the 6 items referred to morning types of activity. The second factor was composed of 3 items (nos. 4, 5, 7) which refer specifically to affect in the morning. The third factor was defined by 3 evening items (nos. 2, 10, 12). The fourth factor was identified by 2 items (nos. 3, 13). The fifth factor had only one item (no. 1) loading above 0.40. Cronbach's alpha values for subscales of the MEQ were factor 1 (alpha=0.809), 2 (alpha=0.735), 3 (alpha=0.623), and 4 (alpha=0.634). The alpha of factor 5 could not be calculated due to its having only a single item. Only factor 1 and factor 2 surpassed the advocated alpha level of 0.70, for the three remaining subscales did not exceed the commonly accepted threshold value of alpha >0.70.

Criterion related validity

Criterion validity coefficients between the two circadian rhythm scales and the criterion variable (dichotomous variable) DSPD diagnosis was eta=0.838 and eta=0.870 for the BALM and MEQ, respectively. Both coefficients are much greater than 0.30, which provides evidence for criterion validity of the two scales. However, when we evaluated the significance of differences between the two correlation coefficients, the two coefficients were not significantly different statistically (Z=1.75, p>0.05).

Classification accuracy

The Area under Curve (AUC) shows the efficiency of the tool: the broader the area is, the more efficient evaluation is. Figure 1 and Table 5 shows the ROC analysis results. The ROC analysis yielded a high AUC estimate of 0.972 for the BALM, and 0.986 for the MEQ.

Further, the AUC values were significant (p<0.001), and the 95% confidence interval (CI) bounds did not include 0.50 (BALM: 95% CI 0.962-0.982; MEQ: 95% CI 0.980-0.992), suggesting diagnostic discrimination that was better than chance alone.

Discriminant analysis

Then we next conducted a discriminant analysis with the BALM to classify the two groups (DSPD vs. normal). In the stepwise solution of the BALM, 6 predictors (item 2, 6, 7, 8, 11, 13) were identified that resulted in good classification accuracy (Table 6 and 7). Examination of the standardized canonical discriminant function coefficients in Table 6 indicates that the best predictor variable that contributed the most to discriminant function, which differentiated uniquely between the DSPD group and normal group, was BALM item 2 (Thinking only of your own "feeling best" times of day, what time would you go to bed if you were completely free to plan your evening?).

Upon examination of the classification procedure, 94.2% of the originally grouped cases were classified correctly by the discriminant function. Indeed, 93.5% of the DSPD group and 95% of the normal group were correctly classified using participants' scores on BALM item 2, 6, 7, 8, 11, 13. In order to cross-validate these results, the leave-one-out classification

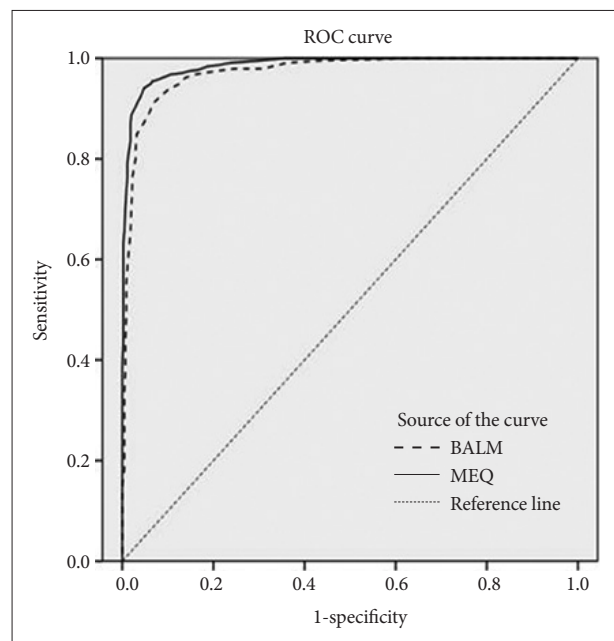


Figure 1. Receiver operating characteristic curve (ROC) for BALM and MEQ. The diagonal line refers to classification level equivalent to a 50% chance occurrence of identifying diagnosed cases. The upper curved line denotes the ROC. BALM: basic language morningness, MEQ: Morningness-Eveningness Questionnaire.

Table 5. Area under the curve

Test result variable(s)	Area	Std. error*	Asymptotic Sig.†	Asymptotic 95% confidence interval	
				Lower bound	Upper bound
BALM	0.972	0.005	<0.001	0.962	0.982
MEQ	0.986	0.003	<0.001	0.980	0.992

*under the nonparametric assumption, †null hypothesis: true area=0.5. BALM: basic language morningness, MEQ: Morningness-Eveningness Questionnaire

Table 6. Results of the discriminant function analysis for the DSPD groups

Predictor	Standardized canonical discriminant function coefficients	Correlations of predictors with standardized canonical discriminant function	Wilks' Lambda	Univariate F (1,879)	Sig.
BALM2	0.419	0.798	0.323	1840.745	<0.001
BALM6	0.140	0.571	0.482	942.788	<0.001
BALM7	0.363	0.768	0.340	1703.720	<0.001
BALM8	0.292	0.655	0.415	1240.427	<0.001
BALM11	0.097	0.537	0.513	833.119	<0.001
BALM13	0.106	0.603	0.456	1048.666	<0.001
Eigenvalue	3.286				
Canonical R	0.876				

Test of function: Wilks' Lambda=0.233, chi-square=1274.886, df=6, p<0.001

DSPD: delayed sleep phase disorder, BALM: basic language morningness

Table 7. Classification of results^{†‡}

		Group	Predicted group membership		Total
			DSPD	NORMAL	
Original	Count	DSPD	415	29	444
		NORMAL	22	416	438
	%	DSPD	93.5	6.5	100.0
		NORMAL	5.0	95.0	100.0
Cross-validation*	Count	DSPD	413	31	444
		NORMAL	23	415	438
	%	DSPD	93.0	7.0	100.0
		NORMAL	5.3	94.7	100.0

*cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case, †94.2% of original grouped cases correctly classified, ‡93.9% of cross-validated grouped cases correctly classified. DSPD: delayed sleep phase disorder

procedure in SPSS was conducted to check the predictive accuracy. Results indicated that 93.9% of the cross-validated group was correctly classified. A discriminant analysis with the MEQ was not done for two reasons, which were 1) BALM has better psychometric properties than the MEQ in many aspects, and 2) nine items of MEQ were overlapped with BALM.

DISCUSSION

Although BALM was similar to MEQ in ROC, it required a smaller number of items and was better than MEQ in item reliability and validity. The main goal of the study was to compare BALM with MEQ in terms of psychometric properties. Comparing internal consistency, Chronbach's coefficient alpha for the BALM (0.898) was higher than that of the MEQ (0.837). Moreover, the corrected item-total correlations were greater than 0.40 in all items of the BALM scale. For the MEQ, seven of the items scored lower than 0.40. This suggests that BALM is superior to MEQ in homogeneity of items. In particular, the α coefficient of 0.898 in BALM is similar to the report by Brown ($\alpha=0.879$ in shift workers, $\alpha=0.91$ in day workers) and hi-

gher than the result by Pornpitakpan ($\alpha=0.781$).^{9,22}

Comparing validity, three factors with eigenvalues of 1.0 or above were extracted from the BALM by using principal axis factoring with varimax rotation, showing high consistency in factor structure between Brown⁹ and the current study done at very different times with very different samples. The alpha values for the subscales of BALM were factor 1 ($\alpha=0.880$), 2 ($\alpha=0.826$), and 3 ($\alpha=0.759$).

The MEQ had considerably lower Cronbach's alpha values, particularly for factors 3, 4, and 5, which had poor consistency. Given the apparent heterogeneity of these 19 items, it is plausible to assume that the items measure different dimensions. These findings suggest that the BALM has better construct validity than the MEQ.

Criterion validity coefficients between the two circadian rhythm scales and the DSPD were $\eta=0.838$ and $\eta=0.870$ for the BALM and MEQ, respectively. Both coefficients are greater than 0.30, which provides evidence for criterion validity of the two scales. The differences in η were not significant, according to Fisher's Z transformation test, though the BALM had slightly better criterion validity.

The efficiency of two assessment tools for DSPD diagnosis was estimated by using the ROC curve nonparametric analysis. The ROC analysis yielded a high AUC estimate of 0.972 in BALM, and 0.986 in MEQ. Two questionnaires AUC estimates were similar and both very good. The MEQ was slightly (but significantly) better at discriminating diagnosis, but this may be partly tautological, since the MEQ had been more influential in arriving at the rater's diagnostic classifications, combined with many other kinds of information integrated in making these diagnostic judgments.

In a discriminant analysis with the BALM to classify the two groups (DSPD vs. normal), 6 items were identified that resulted in good classification accuracy. Upon examination of the classification procedure, 94.2% of the originally grouped cases were classified correctly. Results revealed that 93.5% of the DSPD group and 95% of the normal group were correctly classified using participants' scores on BALM 6 items. Also in cross-validation, 93.9% of the cross-validated group was correctly classified. From these results, we may develop a shortened scale, which consisted of 6 items.

This study provides new evidence that the BALM is a reliable and valid measure of morningness which is psychometrically superior to the MEQ. The present findings indicate BALM scale is effective for distinguishing between diagnosed cases of DSPD and controls judged without DSPD.

Acknowledgments

This research was supported by the U.S. National Institutes of Health's NHLBI HL071123. HJL was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2010-0025130).

REFERENCES

- Guthrie JP. Additional validity evidence for a measure of morningness. *J Appl Psychol* 1995;80:186-190.
- Kerkhof GA. Inter-individual differences in the human circadian system: a review. *Biol Psychol* 1985;20:83-112.
- Horne JA, Östberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Int J Chronobiol* 1976;4:97-110.
- Neubauer AC. Psychometric comparison of two circadian rhythm questionnaires and their relationship with personality. *Pers Individ Diff* 1992;13:125-131.
- Adan A, Almirall H. Standardization of a reduced scale of morningness in Spanish population: individual differences. *Psicothema* 1990; 2:137-149.
- Chelminski I, Petros TV, Plaud JJ, Ferraro FR. Psychometric properties of the reduced Horne and Östberg questionnaire. *Pers Individ Diff* 2000; 29:469-478.
- Greenwood KM. An evaluation of the circadian type questionnaire. *Ergonomics* 1995;38:347-360.
- Smith CS, Reilly C, Midkiff K. Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. *J Appl Psychol* 1989;74:728-738.
- Brown FM. Psychometric equivalence of an improved Basic Language Morningness (BALM) scale using industrial population within comparisons. *Ergonomics* 1993;36:191-197.
- Terman M, Rifkin JB, Jacobs J, White TM. Morningness-Eveningness Questionnaire (Revised). New York, NY: New York State Psychiatric Institute; 2001.
- Schrader H, Bovim G, Sand T. The prevalence of delayed and advanced sleep phase syndromes. *J Sleep Res* 1993;2:51-55.
- Yazaki M, Shirakawa S, Okawa M, Takahashi K. Demography of sleep disturbances associated with circadian rhythm disorders in Japan. *Psychiatry Clin Neurosci* 1999;53:267-268.
- Ando K, Kripke DF, Ancoli-Israel S. Delayed and advanced sleep phase symptoms. *Isr J Psychiatry Relat Sci* 2002;39:11-18.
- Kripke DF, Rex KM, Ancoli-Israel S, Nievergelt CM, Klimecki W, Kelson JR. Delayed sleep phase cases and controls. *J Circadian Rhythms* 2008; 6:6.
- Lee HJ, Rex KM, Nievergelt CM, Kelson JR, Kripke DF. Delayed sleep phase syndrome is related to seasonal affective disorder. *J Affect Disord* 2011;133:573-579.
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6:284-290.
- Field A. *Discovering Statistics Using SPSS*. London: SAGE; 2005.
- Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993; 39:561-577.
- Swets JA. *Signal Detection Theory and ROC Analysis in Psychology and Diagnosis: Collected Papers*. Mahwah, NJ: Erlbaum; 1996.
- Nunnally J. *Psychometrics Theory*. New York, NY: McGraw-Hill; 1978.
- Allen MJ, Yen WM. *Introduction to Measurement Theory*. Long Grove, IL: Waveland Press; 1979.
- Pornpitakpan C. Additional validity of the basic language morningness (BALM) scale. *Pers Individ Diff* 2000;28:59-72.