Reliability and Cross-Validation of the Night Eating Questionnaire (NEQ): Hebrew Version

Yael Latzer, DSc,1,2 Orna Tzischinsky, DSc,3 Michal Hason Rozenstein, PhD,1,2 and Kelly Allison, PhD4

1 Haifa University, Haifa, Israel
2 ED Clinic, Rambam Medical Center, Israel
3 Behavioral Science Department, Emek Yezreel Academic College, Emek Yezreel, Israel
4 University of Pennsylvania School of Medicine, Department of Psychiatry, Philadelphia, Penn., U.S.A.

ABSTRACT

Background: Several developments in diagnosing night eating syndrome (NES) occurred during the last decade. New proposed diagnostic criteria are now available, and a short Night Eating Questionnaire (NEQ) was published. The study aims were to validate the NEQ for assessing NES, to identify the optimal NEQ cut-point for NES diagnosis, and to validate and test the internal reliability of the translated Hebrew version of the NEQ.

Method: 141 participants completed the questionnaire and a diagnostic interview, divided into four groups: NES (n=59), Partial diagnosis PD-NES (n=42), other eating disorders (n=8) and controls (32). Validity was measured by calculating reliability, factor structure, and comparing the interview diagnosis to the NEQ score using different cut scores.

Results: Cronbach’s alpha was 0.78, and principal components analysis yielded a five factor structure. A cut score of 21 provided the best balance of false and true positive diagnosis.

Conclusions: We concluded that the NEQ may be an acceptable screening instrument for assessing NES symptomatology.

INTRODUCTION

Night eating syndrome (NES) was first described in 1955 by Stunkard et al. (1). The criteria for NES have been revised and modified several times since the original description by different authors, mainly in the field of eating disorders (1-6). The core features consistently used to characterize NES since 1955 have been evening hyperphagia (EH), morning anorexia, insomnia and/or other sleep disturbances, nocturnal ingestions (NI), and mood disturbances (3), with evening hyperphagia and nocturnal ingestions representing the core features.

However, none of the modifications seem to have been formally validated, and only one standardized assessment instrument is available. NES was not an official eating disorder diagnosis in the DSM-IV (7), but had been included as a form of other Feeding and Eating Disorders in the DSM-5 (www.dsm5.org).

Variance in the previous definitions of NES has existed across the different features. Evening hyperphagia has been defined differently with regard to the amount of food eaten and the time-frame involved more than 25 percent of daily caloric food intake after the evening meal (8, 9).

Nocturnal ingestion (nocturnal eating) has been defined as waking up at night to eat. It is often associated with feeling a strong urge to eat during the night and/or holding the belief that eating is necessary in order to initiate falling asleep or to continue sleeping (9, 10).

Morning anorexia has been defined with regard to whether breakfast is eaten and/or if hunger is present in the morning (4, 8, 11-13). Finally, sleep disturbances have been described mainly as initial or maintenance insomnia (4, 6, 11, 12), and mood disturbances have included evening tension and/or overall depressed mood, sometimes with mood worsening in the evening (7, 9, 11).
An important differential diagnosis for NES is sleep related eating disorder (SRED) (13, 14). SRED is described as a parasomnia that consists of arising from sleep and eating, with reduced consciousness and partial awareness, reduced level of recall (partial or total amnesia) for night eating, and ingestion of unusual or even non-food items. This has been addressed with the introduction of a proposed set of research criteria for NES (15).

NES DIAGNOSTICS

The need to recognize NES with official diagnostic criteria has become increasingly important. Thus, participants in the First International NES Symposium (April 2008) brought together a set of provisional diagnostic criteria for NES (15). The diagnostic criteria describe a significantly increased caloric intake in the evening and/or night-time, characterized by consuming at least 25% of food intake, and/or experiencing at least two episodes of nocturnal eating per week (criterion A). Awareness and recall of evening and nocturnal eating episodes are present, as a means of differentiating NES from Sleep Related Eating Disorder (criterion B). In addition the daily pattern manifested by at least three of the following features: morning anorexia, a strong urge to eat between dinner and sleep onset and/or during the night, insomnia, a belief that one must eat in order to sleep, and depressed and/or mood worsens in the evening (criterion C). The disorder is associated with significant distress (criterion D), is maintained for at least three months, and is not secondary to medical or psychiatric disorder (criterion E). These proposed criteria currently represent the most updated reference used to diagnose this syndrome.

NES ASSESSMENT

Stunkard and colleagues (6) started developing a questionnaire to screen for the presence of night eating in the 1990s. The first version was a 9-item self-report Night Eating Questionnaire (NEQ) that was not published or validated. Its items assessed: morning anorexia (2 items), evening hyperphagia (1 item), initial insomnia (1 item), mid-phase insomnia (1 item), nocturnal ingestions (1 item), and mood (3 items). Van der Wal and colleagues (16) evaluated the diagnostic utility of various combinations of questions taken from this early version of the NEQ for the recognition of NES. They found that the screening questionnaire had acceptable sensitivity across the various definitions of NES but poor specificity. The authors suggested that NEQ lacked a cut-point that would result in both high sensitivity and specificity and, thus, a better screening tool for NES assessment was needed.

The latest version of the NEQ was validated by Allison et al. in 2008 (17). The NEQ is comprised of 14 items assessing morning hunger and timing of first food consumption (2 items), food cravings and control over eating behavior both before bedtime (2 items) and during night-time awakenings (2 items), percentage of food consumed after dinner (1 item), initial insomnia (1 item), frequency of nocturnal awakenings and ingestion of food (3 items), mood disturbance (2 items), and awareness of nocturnal eating episodes (1 item). This latter item is not included in the total NEQ score as an indicator of severity of NES in the English version, but rather used as a means of differentially diagnosing NES as compared to SRED. Therefore scores range from 0 to 52.

The original English version of the NEQ (17) was shown to have a Cronbach’s alpha of 0.70, and three factors were identified: Nocturnal Ingestions, Evening Hyperphagia and Morning Anorexia. Clinical cut-points were also evaluated, with a cut of 25 yielding a modest positive predictive value of 41% and a higher cut of 30 yielding a positive predictive value of 73%.

The NEQ was translated into Spanish and validated in Spain recently, yielding a Cronbach’s alpha of 0.79 and a similar four-factor solution as the English version (18). Similarly, a Portuguese translation of the NEQ in Brazil yielded a Cronbach’s alpha of 0.78 (19). Finally, the NEQ was translated into French in Canada and adapted for children, using a self-report NEQ and a parent-reported NEQ (20). Cronbach’s alphas were 0.54 for the child version, and 0.55 for the parent-report version.

The only published semi-structured interview for NES is the Night Eating Syndrome History and Inventory (NESHI) (21). It contains questions assessing the schedule and the amount of food eaten in 24 hours in a typical day, history of NES symptoms, sleep patterns, mood and stressors, weight, diet history, and previous treatment strategies. The NEQ items are embedded in the NESHI and can yield a score (17). Its items track onto the research diagnostic criteria so that a diagnosis can be determined.

Given the increasing international use of the NEQ, we sought to translate and evaluate this instrument for use in Hebrew. Our first aim was to validate the NEQ as a brief questionnaire for assessing NES and to identify the optimal NEQ cut-point for NES diagnosis according to the newly proposed diagnostic criteria among an Israeli cohort. Our second aim was to validate and test...
the internal reliability of the translated Hebrew version of the NEQ. In order to evaluate these aims, we compared different classifications of individuals with NES by their NEQ scores, in comparison to an interview-based diagnosis constructed from the 2010 criteria (15) in a community sample.

**METHOD**

**PARTICIPANTS, INSTRUMENTS AND PROCEDURE**

One hundred and forty-one adults between the ages of 18-65 were recruited for the study. All participants were recruited via internet advertisement, as well as from individuals seeking nutritional treatment for weight loss.

We separated the participants into two groups: a clinical group and a control group. The total clinical group consisted of 109 participants who fit one of the two core features of NES according to the 2010 diagnostic criteria. These two features were defined as consumption of at least 25% of intake after the evening meal (evening hyperphagia) and/or nocturnal awakenings with ingestions at least twice per week (nocturnal ingestions) (15).

The clinical group was divided into three subgroups: 1) The NES group (n=59; 44 females, 15 males) met full criteria for NES according to the 2010 criteria; 2) the Partial Diagnosis of NES (PD-NES) group (n=42; 25 females, 17 males) met partial criteria for NES, missing the NES diagnosis by meeting only two instead of three clinical features required in criterion C; and 3) Others (n=8; 7 females, 1 male), who answered “yes” to one of the two main questions, and fulfilled the diagnostic criteria for a different type of eating disorder, rather than one of the two NES subgroups (either bulimia nervosa or binge eating disorder).

The control group consisted of 32 participants (25 females, 7 males) who did not suffer from NES or any other eating or psychiatric disorders.

The study was approved by the Helsinki Committee of Rambam Medical Center and the ethical committee IRB at the University of Haifa.

Prior to undergoing the clinical interview, participants were informed about the purpose of the study and they provided written informed consent. The participants then completed the self-report NEQ. Subsequently, participants were interviewed individually using the NESHI (21), administered by trained clinicians: a clinical social worker, a clinical dietitian, and a clinical psychologist. For the current study, the NESHI was translated into Hebrew, and back-translated into English.

The NEQ was also translated into Hebrew and back-translated into English. The NEQ total score was calculated by reverse coding items 1, 4, and 14, and summing all items (15). The rationale for excluding item 13 (assessing awareness of nocturnal ingestions) in previous studies was that it was used only to differentiate NES cases from SRED cases, but did not contribute to the severity of a particular case of NES. We chose to include the awareness item in the total score in the current study. Participants reporting no awakenings at night and not meeting the criteria for evening hyperphagia were excluded after the interview (due to a sleeping problem rather than eating problem). Thus, the awareness item became relevant for those who do eat at night, so we chose to include it. However, we examined the impact of this decision; our analyses were run with and without item 13 in the total score, with no significant changes in outcomes.

**STATISTICAL ANALYSES**

Descriptive statistics were used to characterize the sample and a histogram was created to show the distribution of NEQ scores for the total sample. Internal consistency was measured using the Cronbach’s alpha procedure. A principal components analysis with Varimax rotation was conducted to extract the factor structure for the Hebrew version of the NEQ. If items loaded above .4 on more than one factor, the higher of the two scores was used if there was a separation of greater than .10. Cross loadings will be discussed.

In order to assess the validity of NEQ, we made a cross validation and compared the outcome of the interview based on the 2010 diagnostic criteria to the NEQ score. First, using Analysis of Variance (ANOVA) we tested whether there were differences in the NEQ scores across the clinical and control groups. Then, using crosstabs (observed vs. expected), we created tables of classification results for three cut-points (i.e., 18, 21, 22). Thus, the scores tested ranged between -0.5 SD and +0.5 SD. We chose to test three different cut-offs that were chosen based on the mean of the entire sample, to see whether they differed from each other in term of true and false positive diagnosis. We also chose them in order to present different scores that can be used for different purposes, such as screening health service or for scientific purposes, which are different in their aims. These cut-scores were used according to the frequency testing the scores that were most frequent (mean + 1 Std) (see Figure 1).

This comparison yields four possible outcomes: 1. True Positive, which results when the NEQ cut-point correctly
identifies people with NES; 2. True Negative, which results when the NEQ identifies people without NES as not having NES; 3. False negative, which results when the NEQ does not identify people with NES as having the syndrome; and 4. False positive, which results when the NEQ mistakenly identifies people without NES as having the syndrome. Theoretically a low cut-score would yield more true positives and false positives, as opposed to a high cut-score, which would yield more false negatives and true negatives. The choice of an optimal cut-score depends on the purpose of diagnosis. However, theoretically, it is important to know which cut-score represents the syndrome at the most accurate rate.

For each cut-score examined, we compared the classification of groups based on the four possible outcomes. We also used Chi-Square tests to compare the groups according to their NEQ scores, and calculated the eta (i.e., the % of the explained variance) to test which cut-score best explained the differences between the groups.

RESULTS

No significant differences were found for age or BMI between groups. In addition no significant differences were found in gender distribution within each group. However, in the Partial Diagnosis of NES (PD-NES) group, there were more males than would be expected compared to the other groups, yet this finding was not statistically significant [Chi-square = 5, p < 0.17]. See Table 1 for groups’ demographic information.

As previous research did not include the awareness item in the total score, it is important to note that the total mean score for all participants using only the 13 items was 21.3, SD = 9.03. The 14 item mean total score was 22.8, SD = 10.3. Means across diagnostic groups for the 14-item scale are presented in Table 1.

FACTOR STRUCTURE

Cronbach’s alpha of the 14-item Hebrew version of the NEQ was 0.78. Principal components analysis with Varimax rotation revealed a 5 factor structure that explained 72% of the scale’s variance. The first factor, Nocturnal Ingestions, explained 25% of the variance and included items 8-12 and 14 (initial insomnia, frequency of awakenings, cravings at night, belief in needing to eat to sleep, frequency of nocturnal eating, and control over nocturnal eating). The second factor, Morning/Evening Eating, explained 14% of the variance and included items 1, 2, 5, and 6 (morning hunger, time of first meal, percentage of food consumed after dinner, and mood). However, item 1 loaded below the .40 level and is considered a weak contributor to the scale. The third factor, Evening Cravings, explained 12% of the variance and included two items, 3 and 4 (cravings and control for eating after dinner). The fourth factor, Awareness, explained 11% of the variance and contained only the awareness item, #13. The fifth factor, Evening Mood, explained 10% of the variance and included items 7 and 8, the initial insomnia item (cross-loaded with Nocturnal Ingestions), and the diurnal variation of mood.

Table 1. Mean (SD) of demographic and NEQ scores described by group

<table>
<thead>
<tr>
<th>Age</th>
<th>BMI</th>
<th>NEQ Score</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean±SD</td>
<td>Range</td>
<td>Mean±SD</td>
<td>Range</td>
</tr>
<tr>
<td>Control</td>
<td>41.64±11.44</td>
<td>26-64</td>
<td>25.14±5.17</td>
</tr>
<tr>
<td>NES</td>
<td>39.94±13.26</td>
<td>19-74</td>
<td>29.88±8.11</td>
</tr>
<tr>
<td>PD-NES</td>
<td>41.74±14.47</td>
<td>18-65</td>
<td>28.69±6.93</td>
</tr>
<tr>
<td>Other</td>
<td>49±10.41</td>
<td>36.5-63</td>
<td>30.41±4.31</td>
</tr>
</tbody>
</table>
EXAMINATION OF CUT SCORES

ANOVA revealed a significant effect by diagnosis [F(3,140) = 33.54, p < .001]. Bonferroni post hoc comparisons revealed significant differences between all group pairs, except between the PD-NES and “Other” group, probably due to the high variance in the “Other” group. This means that the NES group differed from the PD-NES group and the controls, and that the PD-NES group differed from the control, as well. Eta squared was 0.423. As the “Others” group included only eight participants, we excluded them from further analysis.

The second stage of the validation analysis differentiated among the three remaining groups by NEQ cut points (see Table 2 for classification outcomes). Eta was used to investigate the strength of the association between diagnostic group and NEQ cut-scores.

DISCUSSION

The need to recognize and define NES leads to the creation of the newly proposed diagnostic criteria (15). The NEQ is the most widely validated scale for assessing NES (19-21) and can be administered easily as a self-report questionnaire (17).

The purpose of this research was to cross-validate and test the internal consistency of the Hebrew version of the NEQ according to the newly proposed criteria, and identify the optimal NEQ cut-point when screening for the presence of NES. The main results of this yielded high internal consistency at an alpha of 0.78 and a five factor structure. The cut score of 21 provided the best balance of false and true positive diagnosis.

In this first translation and reliability study of the NEQ in Hebrew, the results showed relatively high reliability using the 14 items of the NEQ, which was similar to alphas found in the Spanish (18) and Portuguese (19) translations. A five-item factor structure was extrapolated. The first factor, Nocturnal Ingestions factor, was quite similar to the same factor in the English version (17) except for the addition of the initial insomnia item. This insomnia item cross-loaded on the fifth factor (Evening Mood) which was also where it loaded in the English version. The second factor, Morning/Evening Eating is similar to the Morning Anorexia factor in the English version, while the third factor, Evening Cravings, is similar to the Evening Hyperphagia factor in the English version, with the notable omission of item 5 (percent of calorie intake after dinner) in the Hebrew version. In the English version, item 5 cross loaded on these two factors. In the Hebrew version, it appears on only factor 2. The fourth factor, Awareness, is unique to the Hebrew version. It is interesting that it contains only one item, supporting its use as a tool to differentiate NES from SRED, but not as a symptom that goes hand in hand with other aspects of NES. The final factor, Evening Mood, is similar to the Mood/Sleep factor of the English version.

As for validity, three different cut-scores were tested. The first cut-score of 18 revealed a high level of true positives for NES, the price being a high false positive rate, particularly for the PD-NES group which present similar symptoms of NES but were missing only one of the NES diagnosis criteria.

Therefore a cut-score of 21 may be a more conservative approach to identify NES in community samples using this Hebrew version. When using the NEQ for clinical purposes we would suggest using 18 as a cut-score, because it would misclassify only 6.8% of NES patients. Clinicians may find this cut-point useful since it would also help them identify individuals who are at risk for NES and refer them for further evaluation. On the opposite end of the range of cut-scores, a score of 25 may be useful for research purposes of screening participants for pure NES. Because this is the first study in Israel with the NEQ, we are careful to not endorse a single cut-off point, but instead call for further research.
on the topic. We do, however, suggest using the NEQ as a standard screening self report inventory for dietitians and physicians, who should probably use the lowest cut-off point – 18 in order to refer patients for further diagnosis.

The study results also indicated that the PD-NES group reported lower mean NEQ scores than those of the NES group, and higher than those of the controls, which is in the direction intended.

The resulting cut scores from this study are lower than those reported by Allison et al. (22), despite using all 14 items in the total score in this current study. This difference suggests that caution should be used when attempting to generalize cut-scores for measures across different cultures and with different translations. However, it is in line with Allison et al’s suggestion that different cut-scores may be warranted when using the same measure depending on the purposes and setting.

One of the study’s limitations is the small sample size which did not allow for computation of the different cut-points among those who have nocturnal ingestion, evening hyperphagia, or both. In addition, we are unable to identify which is the best cut-point among people who suffer from NES and who seek treatment for eating disorders, as they were not included in the sample. These limitations call for further research.

In conclusion, the NEQ is a short assessment instrument that can be efficiently administered in a variety of populations and settings and can be used for different purposes using a different cut of point. Although the NEQ was found to be an efficient instrument for identifying NES, for clinical evaluation, we suggest using the Night Eating Syndrome History and Inventory version (NESH) interview for further in-depth diagnosis for those who reach cut-score of 21 or greater. The NESH reviews the typical timing and content of intake across 24-hours.

Similarly, the NEQ total score, which was confirmed to be a valid indicator of a higher-order construct of night eating in this study, may be used for an index of severity, or the factor scales may be examined separately if particular aspects of NES are of interest. Further research should focus on determining its validity and utility for different populations, including eating disorder samples, and subgroups of those with NES.

Reference