JSM Neurological Disorders and Stroke

Research Article

Limitations of Insomnia Severity Index and Possible Remedies

Satyendra Nath Chakrabartty*

Indian Ports Association, Indian Maritime University, India

Abstract

Objectives: To identify major limitations for scoring of Likert type Insomnia Severity Index (ISI) items and to transfer ordinal data to a metric scale satisfying the requirement of measurement theories and to compare empirically the proposed method with usual summative scoring

Methods: Data driven weights assigned to different levels of different items (Stage 1) are followed by choosing weights to different items (Stage 2, before normalization and also after normalization) for meaningful inferences. Properties of the proposed methods discussed. Empirical verification undertaken with hypothetical data.

Results: Generated continuous scores satisfied monotonic, equidistant property with higher discriminating power of the test, facilitated assessment of progress /deterioration of a patient. Further weights to items satisfied additional properties like minimizing variance of test score and/or making the test score equicorrelated with the items, and to find test reliability. Proposed methods passed test of normality. Showed strong linear relationship with the summative score and produced same number of independent factors with different factor loadings.

Conclusion: Resultant Weighted Insomnia Severity scores (WISC) facilitated undertaking analysis in parametric set up. Strong correlations provide the bridge between the psychometric issues in the ordinal /interval controversy. However, considering the theoretical advantages including meaningfulness of operations, better comparison of subjects and parameters of test and items, WISC may be preferred and used for meaningful scoring of ISI items.

Further studies may be undertaken to transform ordinal Likert items to interval or ratio scale exploring examination of unidimensionality of the items comprising a test.

Keywords: Insomnia severity index; Likert items; Weighted sum; Monotonic; Equidistant; Normality

Introduction

Insomnia Severity Index (ISI), a self-reported questionnaire with seven Likert items, each with 5 response categories (Levels) marked as 0,1, 2, 3 and 4, is used to screen and assess degree of insomnia [1]. It has been widely used for research and clinical purposes [2]. The ISI appraises both aspects of insomnia (i.e. nighttime and daytime symptoms [3]. ISI score of a subject is taken as summative score i.e. sum total of numerical values attached to the levels chosen by the subject in each of the 7 items. Minimum and maximum ISI score of an individual are 0 and 28 respectively. A cutoff score of 14 was recommended [4] such that persons scoring less than 14 are to be taken as Normal and those scoring more than 14 to be considered as having insomnia. Such cut-off score have been used by researchers like [5]. The ISI is a useful questionnaire with acceptable validity and reliability for evaluating and screening in the context of primary insomnia [6].

Submitted: 31 July 2018 | **Accepted:** 26 August 2019 | **Published:** 28 August 2019

***Corresponding author:** Satyendra Nath Chakrabartty, Indian Ports Association, Indian Maritime University, N 304, Vivek Vihar, Sector - 82, Noida 201304, India, Tel: 7838-333-652; Email: snc12@rediffmail.com

Citation: Chakrabartty SN (2019) Limitations of Insomnia Severity Index and Possible Remedies. JSM Neurol Disord Stroke 5: 9.

However, self-reported daytime impairment accompanying insomnia does not always match objective evidence of sleep disturbance [7]. People, reporting good sleep, may show disturbed sleep pattern when monitored physiologically. Similarly, people reporting disturbed sleep may show normal sleep patterns when objectively monitored [8]. Analysis involving numerical values attached to the levels (like expected values) are not meaningful because zero is attached to the first level. Absence of method of measuring insomnia intensity of a patient or a group of patients by a continuous variable, may not allow accurate quantification of the progress made by a patient and extent of care needed by an individual patient, impact of a new drug on treatment paradigms and others. Major limitations of ISI are due to use of Likert scale which are associated with limitations and also due to statistical methods used with ordinal ISI data without checking the compliance of the assumptions for such techniques or effect of violation of the assumptions.

The above motivates to identify major limitations for scoring of ISI items (Likert type) along with nature of data generated and to transfer ordinal data to a metric scale that is continuous, objective with a fixed zero point satisfying the requirement of measurement theories and to compare empirically the proposed method with usual summative scoring.

Rest of the paper is organized as follows. Limitations of ISI are discussed in the following section. Section 3 deals with suggested remedies and their properties. Empirical verifications to the suggested methods are described in section 4. The paper is rounded up in Section 5 by recalling the salient outcomes and emerging suggestions.

Limitations of ISI scoring system

Equal importance to the items: The scoring system of ISI, assumes equal importance to all the seven items, despite different values of item-total correlations between 0.47 to 0.75 and inter-item correlation between 0.06 to 0.57 were observed [9]. Principal Component Analysis (PCA) and Factor Analysis (FA) resulted in different number of independent factors with different factor loadings for the items [5,10]. Thus, equal importance to the items may not be justified.

Unequal distance between two successive levels and summative score: Assumes that distance between two successive levels is same for each item. For example, for the Item 4, distance between "Satisfied" (level 1) and "Moderately Satisfied" (level 2) is same as the distance between "Moderately Satisfied" & Dissatisfied" (level 3), which implies distance between "Satisfied" & "Dissatisfied" $(D_{1,3}) = 2D_{1,2} = 2D_{2,3}$ However, ISI scores are unlikely to satisfy such equidistant property. [11]. [12] observed that the distances between each successive response categories of a Likert scale are not equal. Equidistant property of successive levels or response categories $A, B, C, D,$ implies $B - A = C - B = D - C$ for meaningful arithmetic aggregation or $\frac{B}{A} = \frac{C}{B} = \frac{D}{C}$ permitting meaningful computation of geometric mean. If addition is not meaningful then statistics like mean, SD, correlation, analysis like regression, ANOVA, estimation, testing, etc. may not be meaningful [13,14] disfavored using mean and SD with ordinal scales. Cronbach alpha in terms of sum of item variances and test variance could be problematic. Violation of equidistant property may put questions on meaningfulness of operations like addition, taking average, meaningful comparison among respondents and also validity of inferences and use of parametric analysis. However, ISI considers summative score without verification of equidistant property.

Moreover, respondents do not perceive the levels as equidistant, which may distort the interpretations of the test scores [15-17]. Strictly speaking, the numbers assigned to the levels or response categories of a Likert item may not be taken as numbers as such, but only a means to provide ranking responses.

A solution to the problem could be to convert the raw scores emerging from the 7-items by weighted sum where weights are chosen suitably to make the data equidistant.

Presence of zero value to a response category: Presence of zero distorts mean and variance by lowering their values. Similarly, too many zeros to an item will artificially lower the covariance and correlation with that item.

The problems may be avoided if 1 – 5 are assigned to the levels/ response categories, instead of 0 – 4, It will also facilitate meaningful application of mathematical operations. It is well known that nature of generated data remains invariant if the numerical values attached to the levels are replaced by a linear transformation of such numbers.

- **Different ways to get a particular total score:** ISI score does not consider patterns of getting a particular score. Different responses to different items can generate the same aggregate scores for more than one respondent. This may result in: Homogeneous sample with respect to ISI score, but not homogeneous with respect to item scores.
- Zero variance of ISI score for group of subjects with equal score despite variance of item scores.
- Even bipolarized subjects or group of bipolarized subjects would get the same ISI scores.
- If all the subjects endorse a particular response category of an item, variance of that item is zero which implies inter-item correlations involving that item are undefined. Similarly, item-total correlation for that item will be undefined.
- The scale may fail to discriminate among the subjects getting same ISI score.

The above limitations are clarified with a hypothetical data given at Table 1 below where a particular ISI score of seven was obtained in various illustrative ways.

Item reliability: Item-total correlations and inter-item correlations have been reported in many studies on ISI scores without mentioning emerging action to a poor value of inter-item correlation (say 0.06) [9]. Violation of assumptions of product moment correlation like interval or ratio level of measurement, continuous data, normality, etc. may distort item-total correlations

Skewed distribution: Item scores and also test scores of ISI are in general skewed. [3] found mean of ISI score at 17 and SD 5.4 against average score of 21 if each subject prefer to choose level 3, which indicates skewed distribution of ISI score.

Non-verification of assumptions of statistical techniques: Studies with ISI used inferential statistics to find association between insomnia and Health-related quality of life, without verifying assumptions of such techniques [18,3]. Factors measured by Health-related quality of life and ISI could be different. Use of parametric statistics with ordinal data, violation of assumption of normality, etc. were considered as deadly sins [19]. Thus, statistical analysis in the parametric set up cannot be meaningfully undertaken from data set generated from ISI using Likert items. Using parametric analysis, like mean, SD and Pearson's correlation with ordinal data produced very strange results [20]. In addition to testing equidistant property, one needs to have continuous data (unlike discrete scores of Likert items) and test normality and homoscedasticity for applying statistical analysis in parametric set up.

Suggested remedies: To avoid the above said problem areas, new methods of scoring ISI scale is proposed using weighted sum approach where data driven weights are chosen suitably for levels followed by weights for different items ensuring specific magnitude to each level and each item for meaningful inferences, without making any assumptions of continuous nature or linearity or normality for the observed variables or the underlying variable being measured.

Table 1: Various ways to achieve same ISI score.

Subjects	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Total score
1	1	1	1	1	1	0	2	7
2	3	4	0	0	0	0	0	7
3	0	0	0	0	3	0	4	7
4	2	2	3	0	0	0	0	7
5	3	1	1	2	0	0	0	7
6	0	0	0	0	4	0	3	7
7	2	1	0	0	0	0	4	7
Mean	1.57	1.29	0.71	0.43	1.14	0.00	1.86	7.00
variance	1.39	1.63	1.06	0.53	2.41	0.00	2.98	0.00

Observations:

- Score of each subject was 7. But clearly, subject 2 and subject 3 tended to be bipolarized.
- Each subject exhibited different pattern to get same ISI score
- Variance of Item 6 was zero which means correlation of Item 6 with total score is undefined.
- The sample was perfectly homogeneous with respect to ISI score, but not homogeneous with respect to item scores.
- Variance of the items varied despite variance of total of item scores was zero
- Interpretation of same ISI score should be different for different subjects

Possible remedy could be to transform ISI scores of subjects by weighted sum where weights to levels are chosen to satisfy desired propertied including breaking of ties in ISI scores.

Weights to different levels for different items:

Let X_{ij} represents score of the *i*-th individual in the *j*-th item of ISI scale, for $i = 1, 2, ..., n$ and $j = 1, 2, ..., 7$. Thus, X_{ij} takes discrete value between 1 to 5 for 5-levels.

Weighted Insomnia Severity score (WISC) is $W_{ij}X_{ij}$ where $W_{ij}^{'}S$ are different for different levels for different items, and $W_{ij} > 0$, $\sum_{j=1}^{5} W_{ij} = 1$ for each item satisfying the equidistant condition i.e. $W_1, 2W_2, 3W_3, 4W_4, 5W_5$ forms an AP. A positive value of the common difference will ensure satisfaction of monotonic condition (i.e. $5W_5 > 4W_4 > 3W_3 > 2W_2 > W_1$). Steps to arrive at such weights satisfying the above conditions is to derive initial weights ω_{ij} by suitable methods so that $\omega_{ij} > 0$ and $\sum_{j=1}^{5} \omega_{ij} = 1$, followed by correction factor, based on which intermediate weights W_{ij} for *j*= 1, 2, 3...5 are to be calculated and finally selected weights may be obtained by $W_{ij(Final)} = \frac{W_{ij}}{\sum_{j=1}^{5} W_j} = 1$. If common difference is denoted by $\beta > 0$ and $W_1 = \omega_1$, then

$$\beta = 2W_2 - W_1 \Rightarrow W_2 = \frac{\omega_1 + \beta}{2},$$

$$W_3 = \frac{\omega_1 + 2\beta}{3}, W_4 = \frac{\omega_1 + 3\beta}{4} and W_5 \frac{\omega_1 + 4\beta}{5}$$

Note that there could be different method of finding ω_{ij} and β

Proposed Method

In addition to the usual summative method of scoring ISI items with five response categories marked as 1,2,3,4 and 5 (Method 1), the proposed method (Method 2) based on weighted sum approach in two stages is described below:

Method 2.1 (Stage 1- Different weights to levels of different items):

Step 1: For the *i*-th item consider $\omega_{ij} = \frac{f_{ij}}{n}$ for *j*= 1, 2, ... , 5

Clearly, $\omega_{ij} > 0 and \sum_{j=1}^{5} \omega_{ij} = \frac{\sum_{j=1}^{5} f_{ij}}{n} = 1.$

Step 2: Arrange the $W_{ij}^{'}S$ of the *i*-th item in increasing order.

Call them $\omega_{i1}, \omega_{i2}, \omega_{i3}, \omega_{i4} and \omega_{i5}$, where

$\omega_{i1} = \frac{f_{min}}{n} and \omega_{i5} = \frac{f_{max}}{n}$ where maximum and minimum frequency are f_{max} and f_{min} respectively.

Step 3: Consider $W_{i1} = \omega_{i1} = \frac{f_{min}}{n}$.Find the correction factor α so that

$$W_{i1} + 4\alpha = 5W_{i5} \Rightarrow \alpha = \frac{5f_{\max} - f_{\min}}{4n}$$

Define

$$W_{i2} = \frac{\omega_{i1} + \alpha}{2}, W_{i3} = \frac{\omega_{i1} + 2\alpha}{3}; W_{i4} = \frac{\omega_{i1} + 3\alpha}{4}; and W_{i5} = \frac{\omega_{i1} + 4\alpha}{5}$$

Step – 4: Since $\sum_{j=1}^{5} W_{ij}$ is not always equal to one, divide each W by the obtained value of $\sum_{j=1}^{5} W_j$ to get final weights

$$W_{ij(Final)} = \frac{W_{ij}}{\sum_{j=1}^{5} w_j} = 1$$

Here, $j.W_{j(Final)} - (j-1).W_{(j-1)(Final)} = Cons\tan t$

$\forall j = 2,3,4,5$ and $W_{ij(Final)}$ satisfy both monotonic and equidistant conditions.

Observations

i) $W_{ij(Final)}$ are based on empirical probabilities of basic Item-score matrix. The item scores and individual scores are taken as expected values and hence provide measurement of continues variables satisfying conditions of linearity.

ii) If frequency of a particular level of an item is zero, the method may fail and can be taken as zero value for scoring ISI items as weighted sum

iii) Since, $0 < W_{j(Final)} < 1$, mean and variance of test scores as well as item scores will get reduced in comparison to the same from Method 1

iv) Different weights to different levels of different items may break ties of subject scores in Method 1 and thus distinguish the same summative score on the basis of how the score was obtained. The generated scale, thus, increases discriminating power of the test.

v) The method does not guarantee equal contribution of the items. If items contribute differently to test score, addition of scores even after assigning weights to response categories may not be very meaningful.

To avoid the problem, Stage 2 is proposed which takes further weights to items so as to satisfy additional properties like minimizing variance of test score and/or making the test score equicorrelated with the items and explore normality.

Stage 2: Different weights to different items

Method 2.2: Minimum variance of test score

Let X_j be the item score vector resulting from Stage 1 for the *j*-th item, *j*= 1, 2, ... , 7. The vector has *n*-components corresponding to *n*-number of subjects. Following method suggested by [21], it is possible to find the vector of weights $V = (V_1, V_2, ..., V_7)^T$ corresponding to 7 items with $\sum_{j=1}^{7} V_j = 1$ such that variance of the total score (Y) is minimum i.e. to minimize $Var(Y) = V^T DV$ subject to the condition $V^T e = 1$ where D the variance-covariance matrix of the items, *e* is the 7-dimensional vector with each component is equal to 1.

Define $A = V^T DV + \lambda(1 - V^T DV)$ where λ is Lagrangian multiplier. Equating derivatives of A with respect to V and λ to zero, will give $2DV - \lambda e = 0$ and $1 - V^T e = 0$.

Thus, $V = \frac{\lambda}{2} D^{-1} e$ and $V^T e = 1$

Which imply $$V = \frac{D^{-1} e}{e^T D^{-1} e} \qquad (1.1)$$

and $$\lambda = \frac{2}{e^T D^{-1} e} \qquad (1.2)$$

Weights found as above ensures that test score (Y) has minimum variance.

Method 2.3: Equal Item-test correlations:

If the item scores are standardized as $Z_{ij} = \frac{X_{ij} - \overline{X_J}}{S_{X_j}}$ then all the items will be transformed to the same scale and correlation between Y and Z_i is same as the correlation between Y and $Z_j = \frac{1}{\sqrt{e^T R^{-1} e}}$ where R is the correlation matrix and $i \neq j$. In other words, the test score is equi-correlated with the standardized score of each item. In other words, item reliability is same for each item. The method gives positive weights and makes equal importance to the items. Since, standardized item scores are normally distributed for a large sample size and the weighted sum of the standardized score y is likely to satisfy the normality assumption. However, equal correlations may not make the data one-dimensional.

In line with reliability of battery [21], equal item reliability can also be used to find reliability of ISI scale as a function of item reliabilities, as follows:

$$r_{tt(ISI)} = \frac{\sum_{i=1}^{7} r_{tt(i)} S_{Xi} + \sum_{i=1, i\neq j}^{7} \sum_{j=1}^{7} 2COV(X_i, X_j)}{\sum_{i=1}^{7} S_{Xi} + \sum_{i=1, j}^{7} \sum_{j=1}^{7} 2COV(X_i, X_j)} \qquad (1.3)$$

Where $r_{tt(i)}$ and S_{Xi} denote respectively reliability sample SD of the *i*-th item

Advantages of proposed methods:

1. The proposed approach does not make any assumption on distribution of underlying variable or observed variable and avoids most of the limitations of current use of ISI scores.

2. Generates continuous data satisfying equidistant property

3. Has fixed zero point.

4. Like summative score, higher value indicates higher Insomnia Severity

5. Gives data driven different weights to levels of different items and also assigns weights to items

6. Generated scores are monotonic in the sense that choice of *j*-th level will result in higher score than the choice of (*j*-1)-th level for any item for *j*=2, 3, 4, 5

7. Test score is equi-correlated with the standardized score of each item in Method 2.3

8. Proposed approach will help clinicians to take advantages as follows:

Action 1: Instead of the levels marked as 0, 1, 2, 3 and 4, assign numbers 1 – 5 to the levels.

Action 2: Find weights preferably by Method 2.2 or 2.3 and take resultant Weighted Insomnia Severity scores (WISC) as an alternate measure of ISI scores satisfying desired properties from measurement theory angles and take the following advantages:

i. Rank a group of patients with respect to WISC uniquely avoiding ties unlike ranks from usual ISI scores.

ii. Find sample mean and SD for a group of patients.

iii. Assess progress/deterioration of a patient over time points. If X_{it} denotes Weighted Insomnia Severity of the *i*-th patient in *t*-th time period, then $\frac{X_{it} - X_{i(t-1)}}{X_{i(t-1)}} \times 100$ will indicate percentage of progress/deterioration made by the *i*-th patient in *t*-th time in comparison to (*t-1*)-th time period.

iv. Calculate item weights with normalized data to have equal item reliability and estimate reliability of ISI scale avoiding assumptions of Cronbach alpha

Limitations of proposed method

1. Does not involve experimental design behind the data.

2. The method may fail if frequency of a particular level of an item is zero

3. Assumes no missing data.

Empirical verification: For better understanding of the proposed method and to establish feasibility of computations, usual summative score and the proposed method are compared empirically using hypothetical data obtained from administration of 7 Likert items, each with 5- levels to 100 subjects.

Calculation of weights

Weights to different levels for different items (Method 2.1, Stage 1): Shown in Table 2

Weights to items (Stage 2):

After finding the transformed scores for each level of different items, as per Stage 1, weights to Items were found before and also after standardization at Stage 2. Item weights as per Stage 2 are given at Table 3:

Equidistant scores: Difference between successive levels was found to be same for each of the proposed method as can be seen from the Table 4 for Item 1(illustrative)

Thus, continuous scores which are equidistant and monotonic were obtained by each proposed Method. Monotonic property helps to assess progress made or deterioration recorded by a patient. Extent of progress can be evaluated by percentage changes of weighted Insomnia Severity score by any of the proposed method. However, some scores in Stage 2 after standardization became negative.

Tied observations and ranks: Method 1 resulted in large number of tied scores and thus failed to distinguish the respondents with same score. Details are as follows: (Table 5)

No ties were observed for each of Method 2.1, 2.2 and 2.3, considering up to four decimal places. Each subject got unique rank (integer valued) in Method 2.1, 2.2 and 2.3 which remained unchanged across those methods.

Descriptive statistics and Item-total correlations of Method 1 and Method 2: (Table 6)

Reliability:

In (1.3), $S_{X1} = 1, r_{tt(i)} = 0.48 \, and \sum_{i=1, i \neq j}^{7} \sum_{j=1}^{7} 2COV(X_i, X_j) = 4.77532 \forall i.j = 1, 2, ..., 7$

Thus, $r_{tt(ISI)}$ works out to 0.69

Correlations among the methods: Any weighted sum score will be highly correlated with unweighted score since weights indicate change of scale. Different weights to different levels and/or different items will also result in high correlation. Correlations among the methods are shown in Table 7

Factor structure: PCA and FA with Varimax rotation with Kaiser Normalization were undertaken for each method. All the methods resulted in three independent factors with explained cumulative variance between 55.624% to 56.739%. However, factor loadings differed across methods. Rotated Component Matrices for the methods is shown in Table 8

Normality: Anderson – Darling test for Normality was used to

Table 2: Calculation of weights to different levels: Stage 1.

Item	Description	RC-1	RC-2	RC-3	RC-4	RC-5	Total
1	Frequency	3	13	2	31	51	100
	Proportions (ω_{1j})	0.03	0. 13	0.02	0.31	0.51	1.00
	$W_{1j(Final)}$ (α = 0.63)	0.01685	0.18539	0.24157	0.26966	0.28652	1.00
2	Frequency	8	8	4	42	38	100
	Proportions (ω_{2j})	0.08	0.08	0.04	0.42	0.38	1.00
	$W_{2j(Final)}$ (α = 0.505)	0.05146	0.18815	0.23372	0.2565	0.27017	1.00
3	Frequency	9	6	7	39	39	100
	Proportions (ω_{3j})	0.09	0.06	0.07	0.39	0.39	1.00
	$W_{3j(Final)}$ (α = 0.4725)	0.04223	0.18742	0.23581	0.26001	0.27453	1.00
4	Frequency	3	2	7	31	57	100
	Proportions(ω_{4j})	0.03	0.02	0.07	0.31	0.57	1.00
	$W_{4j(Final)}(\alpha = 0.7075)$	0.01016	0.18486	0.24309	0.27221	0.28968	1.00
5	Frequency	4	8	6	47	35	100
	Proportions(ω_{5j})	0.04	0.08	0.06	0.47	0.35	1.00
	$W_{5j(Final)}$ (α = 0.5775)	0.02409	0.18597	0.23993	0.26691	0.28310	1.00
6	Frequency	5	6	4	38	47	100
	Proportions (ω_{6j})	0.05	0.06	0.04	0.38	0.47	1.00
	$W_{6j(Final)}$ (α = 0.5775)	0.02971	0.18642	0.23865	0.26477	0.28045	1.00
7	Frequency	8	4	4	38	46	100
	Proportions(ω_{7j})	0.08	0.04	0.04	0.38	0. 46	1.00
	$W_{7j(Final)}$ (α = 0.565)	0.02460	0.18601	0.23982	0.26672	0.28286	1.00

Legend: RC- j ⟹ j-th Response category for j=1,2,3,4,5

Table 3: Item weights: Stage 2.

Item No.	Method 2.2	Method 2.3
Item 1	0.126357	0.159962
Item 2	0.060899	0.071872
Item 3	0.150605	0.167070
Item 4	0.173187	0.102129
Item 5	0.206746	0.178654
Item 6	0.171362	0.152726
Item 7	0.110843	0.167587
Sum of weights	1.00	1.00

Table 4: Equidistant scores of Item 1 of proposed Methods.

Scores of Item 1	1 in Method 1	2 in Method 1	3 in Method 1	4 in Method 1	5 in Method 1
Transformed score in Method 2.1	0.01685	2(0.18539)= 0.37078	3(0.24157)= 0.72471	4(0.26966)=1.07864	5(0.28652)=1.4326
Difference between successive levels		0.3539 (2 & 1)	0.3539 (3 & 2)	0.3539 (4 & 3)	0.3539 (5 & 4)
Transformed score in Method 2.2	0.00213	0.04685	0.09157	0.13629	0.18102
Difference between successive levels		0.04472 (2 & 1)	0.04472 (3 & 2)	0.04472 (4 & 3)	0.04472 (5 & 4)
Transformed score in Method 2.3	-0.42449	-0.28969	-0.15489	-0.02009	0.11479
Difference between successive levels		0.1348 (2 & 1)	0.1348 (3 & 2)	0.1348 (4 & 3)	0.1348 (5 & 4)

Table 5: Number of tied score in Method 1.			
Method 1			
Score	**No. of ties**	**Score**	**No. of ties**
17	2	28	9
22	3	29	7
23	2	30	17
24	5	31	13
25	4	32	16
26	5	33	5
27	7	34	3

Table 6: Descriptive statistics and Item-total correlations.

Description	**Method 1 (Summative score)**	**Method 2.1 (Weights to levels for different items)**	**Method 2.2 (Weights to items before standardization)**	**Method 2.3 (Weights to items after standardization)**
Test Mean	28.65	7.69373	1.107085	-0.00373
Test variance	14.85606	1.82497	0.035671	0.232708
Range of inter-item correlations	(-)0.03 to 0.30	(-)0.02 to 0.25	(-)0.02 to 0.25	(-)0.02 to 0.25
No. of negative inter-item correlations	3	1	1	1
Item-total correlations				
Item 1	0.44795	0.48165	0.450116	0.483433
Item 2	0.61774	0.59687	0.482573	0.485325
Item 3	0.48364	0.46773	0.461145	0.484177
Item 4	0.57860	0.52654	0.568488	0.484583
Item 5	0.39051	0.41460	0.517787	0.483776
Item 6	0.46102	0.47316	0.496722	0.480124
Item 7	0.46449	0.47185	0.412237	0.483543

Observations:

- Weighted sum reduced mean and variance of test score. Test variance at the level of 0.03567 was minimum for Method 2.2 (without standardization).
- The items were equi-correlated with test score for Method 2.3 (after standardization) (up to two decimal places).
- Number of negative inter-item correlations was maximum for Method 1
- Further weights to items were taken in Method 2.2 to minimize test variance (before standardization) and Method 2.3 to make the test score equicorrelated with the item scores (after standardization).

Table 7: Correlations among the methods.

	Method 1	**Method 2.1**	**Method 2.2**	**Method 2.3**
Method 1	1.00	0.97782	0.96215	0.95791
Method 2.1		1.00	0.97959	0.98730
Method 2.2			1.00	0.99149
Method 2.3				1.00

Highest correlation was found between Method 2.2 and 2.3. Strong linear relationship of weighted sum scores with the summative score (Method 1) is likely to give similar factor structure except the values of factor loadings and some inter-item correlations.

test H_0 **:** Test scores follow Normal distribution. H_0**:** was accepted for each method except Method 1.

Conclusion

Weighted sum approach of scoring ISI scale was proposed where weights are first assigned to different levels of different items (Stage 1) followed by choosing weights to different items (Stage 2, before normalization and also after normalization). The approach used only the frequencies of levels without making any assumptions of continuous nature or linearity or normality for the observed variables or the underlying variable being measured. Resultant Weighted Insomnia Severity scores (WISC) were continuous, satisfied monotonic and equidistant properties, made the data homogenous, passed normality test, avoided tied

Table 8: Rotated Component Matrix for the methods.

	Factor 1	Factor 2	Factor 3
Item 1	0.152 (**0.657**) **[0.659]** **[[0.657]]**	-0.132 (0.098) [0.097] [[0.098]]	**0.779** (-0.087) [-0.091] [[-0.087]]
Item 2	**0.669** (0.313) [0.311] [[0.313]]	0.141 **(0.650)** **[0.651]** **[[0.650]]**	0.329 (0.150) [0.150] [[0.150]]
Item 3	**0.648** (-0.138) [-0.139] [[-0.138]]	-0.037 (0.407) [0.407] [[0.407]]	0.057 **(0.620)** **[0.619]** **[[0.620]]**
Item 4	0.242 **(0.504)** **[0.503]** **[[0.504]]**	**0.695** (0.433) [0.434] [[0.433]]	0.239 (0.055) [0.055] [[0.055]]
Item 5	**0.618** (-0.092) [-0.093] [[-0.092]]	0.163 **(0.734)** **[0.734]** **[[0.734]]**	-0.292 (-0.062) [-0.062] [[-0.062]]
Item 6	-0.106 **(0.757)** **[0.756]** **[[0.757]]**	0.345 (-0.089) [-0.087] [[-0.089]]	**0.646** (0.123) [0.125] [[0.123]]
Item 7	-0.010 (0.135) [0.134] [[0.135]]	**0.820** (-0.136) [-0.135] [[-0.136]]	-0.105 **(0.840)** **[0.840]** **[[0.840]]**

Note: Figures without parentheses denote factor loadings by Method 1; figures within ( ) denote factor loadings by Method 2.1; figures within [ ] denote factor loadings by Method 2.2 and figures within [[ ]] denote same for Method 2.3

Observations:

- Factor loadings by Method 1 differed with those obtained from Method 2
- Factor loadings by Method 2.1, 2.2 & 2.3 were more or less same and thus there is not much difference in exploring and interpreting factors.
- Maximum loadings by Method 2 : Factor 1 – Item 1, 4 and 6
  Factor 2 – Item 2 and 5
  Factor 3 – Item 3 and 7
- Extraction of three factors for each method clearly indicates that the scores are not one-dimensional and hence Coefficient alpha may not be advisable. Test reliability was found using (1.3).

scores and facilitated assessment of progress/deterioration of a patient and undertaking analysis in parametric set up. Each generated metric using weighted sum had fixed zero value. Each subject got unique rank (integer valued) in Method 2.1, 2.2 and 2.3 which remained unchanged across those methods. Weights to items had additional property of minimum test variance when weights are selected before normalization. However, item weights calculated with normalized data resulted in equal item reliability i.e. all items were eqicorrelated with test score, which also helped to estimate reliability of test.

The proposed methods showed strong linear relationship with the summative score method (Method 1). Highest correlation was found between Method 2.2 and 2.3. Strong linear relationship of weighted sum scores with the summative score produced same number of independent factors though factor loadings were different from the same obtained from summative scores.

High correlations among the methods may reconcile the debate between the ordinals - interval controversy, in the sense that there may not be much harm in using data generated from summative scores of Likert questionnaire. High linear relationship between weighted sum scores and unweighted score may be accepted by those who favored treating Likert items as interval and those who oppose such idea.

However, considering the theoretical advantages of weighted sum approaches including meaningfulness of operations, better comparison of subjects and parameters of test and items, platform to undertake parametric statistical analysis, the proposed Method 2.2 or 2.3 may be preferred and used for scoring ISI items.

Simulation studies may be undertaken to study the above said proposed measures in multi data sets. Further studies may be undertaken to explore non-linear approaches to transform ordinal Likert items to interval or ratio scale exploring examination of unidimensionality of the items comprising a test.

References

1. Bastien, CH, Vallieres A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Medicine. 2001; 2: 297-307.

2. Morin CM. *Treatment manuals for practitioners. Insomnia: Psychological assessment and management. 1993;* New York, NY, US: Guilford Press.

3. Bluestein D, Rutledge CM, Healey AC. Psychosocial Correlates of Insomnia Severity in Primary Care. J Am Board Fam Med. 2010; 23: 204-211.

4. Kaczor M, Skalski M. Prevalence and consequences of insomnia in pediatric population. Psychiatr Pol. 2016; 50: 555–569.

5. Chahoud M, Chahine R, Salameh P, Sauleau EA. Reliability, factor analysis and internal consistency calculation of the Insomnia Severity Index (ISI) in French and in English among Lebanese adolescents. eNeurologicalSci. 2017; 7: 9-14.

6. Savard MH, Savard J, Simard S, Ivers H. Empirical validation of the Insomnia Severity Index in cancer patients. Psych-oncology. 2005; 14: 429-441.

7. Orff HJ, Drummond SP, Nowakowski S, Perlis ML. Discrepancy between subjective symptomatology and objective neuropsychological performance in insomnia. Sleep. 2007; 30: 1205-1211.

8. Edinger JD, Fins AI, Glenn DM, Sullivan RJ Jr, Bastian LA, Marsh GR, et al. Insomnia and the eye of the beholder: Are there clinical markers of objective sleep disturbances among adults with and without insomnia complaints? J Consult Clin Psychol. 2000; 68: 586-593.

9. Manzar MD, Salahuddin M, Khan TA, Shah SA, Alamri M, Pandi-Perumal SR, et al. Psychometric properties of the Insomnia Severity Index in Ethiopian adults with substance use problems. J Ethn Subst Abuse. 2018; 1-15.

10. Osorio RS, Pirraglia E, Agüera-Ortiz LF, During EH, Sacks H, Ayappa I, et al. Greater risk of Alzheimer's disease in older adults with insomnia. J Am Geriatr Soc. 2011; 59: 559-562.

11. Ferrando PJ. A Kernel density analysis of continuous typical-response scales. Educational and Psychological Measurement. 2003; 63: 809-824.

12. Munshi J. A method for constructing Likert scales. 1990.

13. Stevens SS. Mathematics, measurement and Psychophysics. In Handbook of Experimental Psychology. SS Stevens (edn), 1951. New York: John Wiley & Sons. 1-49.

14. Jamieson S. Likert scales: how to (ab) use them. Med Educ. 2004; 38: 1212-1218.

15. Lee JA, Soutar GN. Is Schwartz's value survey an interval scale, and does it really matter? Journal of Cross-Cultural Psychology, 2010; 41: 76-86.

16. Lantz B. Equidistance of Likert –Type Scales and Validation of Inferential Methods using Experiments and Simulation. EJBRM. 2013; 11: 16-28.

17. Mundy J, Dickinson D. Factors affecting the uptake of voluntary HIV/AIDS counseling and testing (VCT) services in the workplace. In: HIV/AIDS in the Workplace Research Symposium. 2004; 175-193.

18. Katz DA, McHorney CA. The Relationship Between Insomnia and Health-Related Quality of Life in Patients With Chronic Illness. J Fam Pract. 2002; 51: 229-235.

19. Kuzon WM, Urbanchek MG, McCabe S. The seven deadly sins of statistical analysis. Ann Plast Surg. 1996; 37: 265-272.

20. Marcus-Roberts HM, Roberts FS. Meaningless statistics. Journal of Educational Statistics, 1987; 12: 383-394.

21. Chakrabartty SN, Wang K, Chakraborti D. Estimation of Uncertainty of Test Score Data using a New, Fast Method of Number Partitioning, Research Gate. 2015.