ROMANIAN JOURNAL
OF PSYCHOLOGICAL STUDIES
HYPERION UNIVERSITY
www.hyperion.ro

TRANSFORMING LIKERT SCORES TO RATIO SCALE

SATYENDRA NATH CHAKRABARTTY
Indian Ports Association, Indian Maritime University

Abstract

Two methods of scoring Likert items as weighted sum are proposed with different weights to response categories of different items using only the frequencies of Item–Response categories. The methods result in continuous data satisfying equidistant property with specified zero value and thus provide platform to perform parametric analysis. Proposed approaches made the data homogenous and distinguished the same summative score on the basis of how the score was obtained. The proposed methods and the summative score method passed normality test and produced same number of independent factors with different factor loadings. Reconciliation of the ordinal - interval controversy is reached in the sense that there may not be much harm in using summative Likert scores which correlates high with the weighted sum approaches. However, considering the theoretical advantages including meaningfulness of operations and comparisons, platform to undertake parametric statistical analysis and easiness to compute weights, no tied scores and no outliers, method based on area under the N (0, 1) may be used for making equidistant Likert scores.

Keywords: *Likert items, weighted sum, monotonic and equidistant, normal curve, ordinal-interval*

1. INTRODUCTION

Likert scales are used frequently in social science, health status, and survey. However, concern is raised on measurement aspects of ordinal data with unknown distance between successive categories, generated by such a scale and associated operations to permit parametric analysis and hence validity of outcomes. Lantz, (2013) observed that the subjects do not perceive Likert type scale as equidistant and suggested further studies on subject perceptions for such scale. Predominance of articles treating ordinal data as interval was observed by Harwell & Gatti, (2001).

Large volume of literature exists on controversy between treating ordinal data as interval and use of parametric analysis (Wilson, 1971; Gaito, 1980; Townsend & Ashby, 1984; Narens & Luce, 1986; Velleman & Wilkinson, 1993; Zumbo &

Corresponding author: Satyendra Nath Chakrabartty
E-mail address: snc12@rediffmail.com

Zimmerman, 1993; Hand, 1996; Jamisson, 2004). Major reasons of such ordinal – interval controversy appears to be non- availability of a globally accepted transformation of Likert scores satisfying desired properties from measurement theory point of view. Wu and Leung, (2017) suggested increasing the number of Likert scale points to 11 to bring the scale closer to the underlying distributions with lower values of Kolmogorov-Smirnov (KS) statistics and interval scale. However, increasing Likert scale points may aggravate the problem of non-equidistance in Likert items.

The problem of ordinal data has been addressed by researchers in terms of rescaling ordinal data to scales having properties of interval level measurement so that parametric statistics can be used. The following important approaches merit consideration:

i) Alternative least squares optimal scaling (ALSOS): A model driven technique by Jacoby (1999) to provide optimal set of measurement values assuming that the respondents construe questions in a similar way. The assumption is generally not found to be true and failure to correctly specify the model will result in biased values of the optimally scaled variables generated from iterative process.

ii) Item response theory (IRT): It is theoretically sound and needs satisfaction of rigorous assumptions. Further explorations may be made for Likert scale with no "correct" or "wrong" answer, Granberg-Rademacker (2010). Harwell and Gatti (2001) used IRT to rescale scores from 5-point rating scale into intervals and found that the model did not adequately fit the data. From IRT, even large ordinal scales can be radically non-linear.

iii) Anchoring vignettes (AVs): Advocated by King et al. (2003) attempts using AVs to better mitigate the problem of differential item functioning (DIF). AVs are application-specific remedies that ask respondents to answer questions about hypothetical aspects described in the vignettes. AVs offer a strong theoretical and practical way of handling DIF, but have practical difficulties like need to have more than one vignette for a given latent attitude or variable and vignette responses are not available in secondary datasets.

iv) Markov Chain Monte Carlo Scaling (MCMC): By Granberg-Rademacker (2010) is a method based on a multivariate normal distribution and Bayesian methodology to find prior information about the latent, unobserved variable to calculate the posterior distribution, requiring a number of iterations. Empirically, MCMC performed better than OLS, ALSOS monotonic, and ALSOS non-monotonic approaches.

In addition to complex procedure involved in each of the above said method of conversion, question arises on how accurate the rescaled data represent the actual data. Score of a subject in Likert scale is obtained by adding his/her score in each item, presuming that the items have equal weights. But a particular score may be obtained by several subjects in several fashions. Thus, interpretation of score of a subject needs to consider how the score was obtained i.e. to consider the response categories of the items endorsed by the subject. Moreover, results of Factor Analysis (FA) usually revels that factor loadings of the items comprising that scale differ and thus contradicts justification of equal weights to the items. If Principal Component Analysis (PCA) is conducted, sum of factor loadings is in general different from one. FA or PCA attempts to assign different weights to different items. Attempts to have different weights to different response categories of different items to score Likert items are rather rare. Barua (2013) used weights to each Likert item of the questionnaire considering item score, weights by Content Experts, internal reliability or Cronbach's alpha and Item discrimination index (IDI) in terms of Spearman's Correlation. Major limitations of this approach are (i) Item score assumes equal weights to items, (ii) Weights by Content Experts are subjective. Reliability or agreement among the experts needs to be considered also (iii) IDI as Spearman correlation with total score and Item reliability as correlation between item score and total score - are questionable. If all subjects choose a particular response category of an item, variance of the item is zero and thus correlation with total score or with any other item is undefined. Cronbach's α may not perform well for discrete data with more than one dimension. Zijlmans et al. (2018) observed that Cronbach's α cannot be used to estimate item-score reliability, (iv) There are other methods with different sets of assumptions to measures Item discrimination index. Suitable measure of weights to items can be attempted where weights are not calculated using sum of Likert scores since addition may not be meaningful.

Thus, need is felt to have new methods of scoring Likert items as weighted sum where weights are assigned to different response categories using only the frequencies or empirical probabilities of Item – Response categories to have continuous data satisfying desired properties and providing a platform to perform analysis parametric set up and to compare the methods and also to have reconciliation of debate on ordinal vs. interval level of Likert scoring in terms of correlation between proposed weighted sum approaches and usual summative scores.

The rest of the paper is structured as follows. In the following Section, the methodology for obtaining weights to response categories of Likert items is presented. Section 3 elaborates on such weights and properties of such weighted sums. Empirical verification for the proposed methods is discussed in Section 4. The paper is rounded up in Section 5 by recalling the salient outcomes of the work and suggesting reconciliation of debate on ordinal vs. interval level.

2. METHODOLOGY

The starting point is the basic Item- score matrix $((X_{ij}))$ where *n*-individuals are in rows and scores of *m*- Likert items are in columns and X_{ij} represents score of the *i*-th individual in the *j*-th item and takes discrete value between 1 to 5 for a 5-point scale.

For the usual summative scoring method, $\sum_{j=1}^{m} X_{ij}$ gives score of the *i*-th individual; $\sum_{i=1}^{n} X_{ij}$ indicates score of the *j*-th item and total test score i.e. sum of scores of all the individuals on all the items is obtained by $\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}$.

Instead of assigning weights to items, it may be prudent to assign different weights to different response categories for different items, where the weights are positive and satisfy $\sum_{j=1}^{5} W_{ij} = 1.$ The transformed score of the *i*-th subject for choosing the *j*-th response category of an item is$W_{ij}X_{ij}$. Thus, both individual scores and item scores provide measurements of continuous variable. However, the transformed scores as weighted sum should satisfy the equidistant property i.e. $5W_5 - 4W_4 = 4W_4 - 3W_3 = 3W_3 - 2W_2 = 2W_2 - W_1$. The condition is satisfied if and only if $W_1, 2W_2, 3W_3, 4W_4, 5W_5$ forms an Arithmetic Progression (AP). A positive common difference ensures that for an item, if a subject chooses *j*- th response category (say 4), his/her transformed score for the item must be greater that the transformed score if he/she had chosen *(j-1)*-th response category (say 3).

To ensure satisfaction of the equidistant property, it is proposed 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 $\alpha > 0$ and $W_1 = \omega_1$, then $\alpha = 2W_2 - W_1 \Longrightarrow W_2 = \frac{\omega_1+\alpha}{2}, W_3 = \frac{\omega_1+2\alpha}{3}, W_4 = \frac{\omega_1+3\alpha}{4}$ and $W_5 = \frac{\omega_1+4\alpha}{5}$

Note that

i) The above said method of finding common difference holds irrespective of process of finding initial weights so long the initial weights add up to one.

ii) However, a negative α may not satisfy $W_2 < W_1$ and $2W_2$ may not exceed W_1.

iii) If weights are based on empirical probabilities of basic Item score matrix, then item scores and individual scores are obtained as expected values and hence provide measurement of continues variables satisfying conditions of linearity.

iv) The metric data and linearity of scores by weighted sum by the above said approach enables generation of scores that is cardinal, equidistant and continuous to permit calculation of almost all descriptive statistics and also to undertake relevant estimation, testing of hypothesis, relevant analysis used in multivariate statistics. However, normality needs to be tested empirically.

v) If frequency of a particular response category of an item is zero, the method may fail and can be taken as zero value for scoring Likert items as weighted sum

3. PROPOSED METHODS

In addition to the usual summative method of scoring Likert items with five response categories marked as 1,2,3,4 and 5 (Method 1), two proposed methods are described below.

3.1 METHOD 2 (BASED ON FREQUENCY OF EACH RESPONSE CATEGORY)

Here, initial weight to *j*-th response category of the *i*-th item is defined as the ratio of frequency of the *j*-th response category of the *i*-th item and total number of subjects i.e. $\omega_{ij} = \frac{f_{ij}}{n}$.

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

Observations:

- Initial weight (ω_{ij}) depends heavily on frequency of the response category i.e. f_{ij}
- $\omega_{ij}'s$ may not follow increasing order
- The initial weights do not satisfy the monotonic condition. For example, if $f_{i5} < f_{i4}$ then $5\omega_{i5}$ may not exceed $4\omega_{i4}$.
- Weighted sum score based on ω_{ij} do not satisfy the equidistant condition

CALCULATION OF FINAL WEIGHTS FOR METHOD 2:

To satisfy the monotonic and the equidistant conditions, correction factor is required based on which intermediate weights W_{ij} and finally selected weights $W_{ij(Final)}$ for j= 1, 2, 3...5 are to be calculated. The suggested steps are as follows:

Step-1: Arrange the $\omega_{ij}'s$ of the *i*-th item in increasing order.

Call them ω_{i1} ω_{i2}, ω_{i3}, ω_{i4} *and* ω_{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 – 2: To ensure that the transformed scores satisfy the equidistant property, 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 – 3: However, $\sum_{j=1}^{5} W_j \neq 1$. To make sum of the weights equals to one, divide each W by the obtained value of $\sum_{j=1}^{5} W_j$ and get final weights $W_{ij(Final)} = \frac{W_{ij}}{\sum_{j=1}^{5} W_j}$

Here, $j.W_{j(Final)} - (j-1).W_{(j-1)(Final)} = \frac{\alpha}{\sum_{j=1}^{5} W_j} \forall j = 2,3,4,5$

Thus, from the initial weights ω_{ij}'s the final weights $W_{ij(Final)}$ can be obtained as per the correction factor α and satisfying equidistant property.

3.2 METHOD 3(BASED ON AREA UNDER STANDARD NORMAL CURVE)

Initial weights based on area under the $N(0, 1)$ are obtained as follows:

Step 1: For the i-th item and the j-th response category, find proportion of frequency of the response category to the sample size i.e. $p_{ij} = \frac{f_{ij}}{n}$. Also find the cumulative proportions (C_i)

Step 2: Find area (A_i) under the standard Normal curve for eachC_i.

Step 3: Take initial weights as $\omega_{ij} = \frac{A_i}{\sum A_i}$

Observations:

1. Sum of initial weights is equal to one i.e. $\sum_{j=1}^{5} \omega_{ij} = 1$

2. The initial weights are in increasing order and thus satisfy the monotonic condition (1.1) under the assumption of $p_{ij} > 0$.

CALCULATION OF FINAL WEIGHTS FOR METHOD 3:

To make the transformed scores equidistant for a 5-point scale, consider the correction factor $\gamma = \frac{Max.area - Min.area}{3}$ since the difference between Maximum area and the Minimum area is contributed by 3 response categories. Determine the modified areas $\Delta_1, \Delta_2, \Delta_3, \Delta_4$ and Δ_5 as follows:

$\Delta_1 = A_1$(unchanged), $\Delta_2 = \frac{\Delta_1+\gamma}{2}$, $\Delta_3 = \frac{\Delta_1+2\gamma}{3}$, $\Delta_4 = \frac{\Delta_1+3\gamma}{4}$, $\Delta_5 = \frac{\Delta_1+4\gamma}{5}$

Since $\sum_{j=1}^{5} \Delta_j \neq 1$, define final weights $W_{j(Final)} = \frac{\Delta_j}{\sum_{j=1}^{5} \Delta_j}$

Here, $j.W_{j(Final)} - (j-1).W_{(j-1)(Final)} = \frac{\gamma}{\sum_{j=1}^{5} \Delta_j} \forall j = 2,3,4,5$

The transformed scores based on the final weights so defined ensure equidistant condition and also satisfy $\sum_{j=1}^{5} W_{j(Final)} = 1$,

Summary:

In Method 2 and 3:

- Initial weights consider empirical probabilities obtained from data considering the frequencies of Item – Response categories without involving assumptions of continuous nature or linearity or normality for the observed variables or the underlying variable being measured.

- Each $W_{j(Final)} > 0$.
- Score emerging from each of the proposed weighted sum approaches satisfy equidistant property and also satisfy $\sum_{j=1}^{5} W_{j(Final)} = 1$
- Item and individual scores are obtained as expected values and generate cardinal data.
- Since, $0 < W_{j(Final)} < 1$, mean and variance of test scores as well as item scores from Method 2 and Method 3 will get reduced in comparison to the same from Method 1.

4. EMPIRICAL VERIFICATION

The method of usual summative score and the two proposed methods are compared empirically using data obtained from administration of a 5- point Likert test with 7 items to 101 subjects, who could answer completely the entire questionnaire.

4.1 CALCULATION OF WEIGHTS

Calculation of weights to various response categories of different items by Method 2 and Method 3 are given at Table 1 and Table 2 respectively.

Table 1

Calculation of weights to response categories of Items: Method 2

Item	**Description**	**RC-1**	**RC-2**	**RC-3**	**RC-4**	**RC-5**	**Total**
1	Frequency	4	9	3	36	49	101
	Proportions(ω_{1j})	0.03960	0.0891	0.02970	0.35644	0.48515	1.00
	Intermediate weights(W_{1j}) (α = 0.599)	0.03960	0.31930	0.41253	0.45915	0.48712	1.717
	Final weights ($W_{1j(Final)}$)	0.02306	0.18589	0.24016	0.26730	0.28359	1.00
2	Frequency	49	25	6	13	8	101
	Proportions (ω_{2j})	0.48515	0.24753	0.05941	0.12871	0.07921	1.00
	Intermediate weights(W_{2j}) α =0.59158)	0.48515	0.53837	0.55611	0.56497	0.57030	2.714S4
	Final weights ($W_{2j(Final)}$)	0.17870	0.19830	0.20483	0.20810	0.21006	1.00
3	Frequency	27	32	10	18	14	101
	Proportions(ω_{3j})	0.26733	0.31683	0.09901	0.17822	0.13861	1.00

	Intermediate weights(W_{3j}) (α =0.37129)	0.26733	0.31931	0.33663	0.34530	0.35049	1.6196
	Final weights ($W_{3j(Final)}$)	0.16511	0.19722	0.20792	0.21327	0.21648	1.00
4	Frequency	5	12	11	31	42	101
	Proportions(ω_{4j})	0.04950	0.11881	0.10891	0.30693	0.41584	
	Intermediate weights(W_{4j}) (α =0.50743)	0.04950	0.27846	0.35478	0.39295	0.41584	1.4913
	Final weights ($W_{4j(Final)}$)	0.03319	0.18670	0.23786	0.26345	0.2788	1.00
5	Frequency	6	13	7	33	42	101
	Proportions(ω_{5j})	0.05941	0.12871	0.06931	0.32673	0.41584	1.00
	Intermediate weights(W_{5j}) (α =0.50495)	0.05941	0.28218	0.35644	0.39356	0.41584	1.5076
	Final weights ($W_{5j(Final)}$)	0.03941	0.18719	0.23645	0.26108	0.27586	1.00
6	Frequency	29	14	12	25	21	101
	Proportions(ω_{6j})	0.28713	0.13861	0.11881	0.24752	0.20792	1.00
	Intermediate weights(W_{6j}) (α =0.329208)	0.28713	0.30817	0.31518	0.31869	0.32079	1.549
	Final weights ($W_{6j(Final)}$)	0.18525	0.19882	0.20335	0.20561	0.20697	1.00
7	Frequency	9	14	6	32	40	101
	Proportions(ω_{7j})	0.08911	0.13861	0.05941	0.31683	0.39604	1.00
	Intermediate weights(W_{7j}) (α =0.480198)	0.08911	0.28465	0.34983	0.38243	0.40198	1.5083
	Final weights ($W_{7j(Final)}$)	0.05909	0.18876	0.23199	0.25360	0.26657	

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

Table 2

Item	**Description**	**RC-1**	**RC-2**	**RC-3**	**RC-4**	**RC-5**	**Total**
1	Cumulative Proportions	0.03960	0.12871	0.15842	0.51485	1.00	
	Area under N(0,1)	0.516	0.5517	0.5636	0.695	0.8413	

	Modified area(Δ_1) (γ =0.108433)	0.516	0.31222	0.24429	0.21033	0.18995	1.47278
	Final weights ($W_{1j(Final)}$)	0.35036	0.21199	0.16587	0.14281	0.12897	1.00
2	Cumulative Proportions	0.48515	0.73267	0.79208	0.92079	1.00	
	Area under N(0,1)	0.6879	0.7673	0.7852	0.8212	0.8413	
	Modified area(Δ_2) (γ =0.051133)	0.6879	0.36952	0.26339	0.21032	0.17849	1.70962
	Final weights ($W_{2j(Final)}$)	0.40237	0.21614	0.15406	0.12302	0.10440	1.00
3	Cumulative Proportions	0.26733	0.58416	0.68317	0.86139	1.00	
	Area under N(0,1)	0.6064	0.7190	0.7517	0.8051	0.8413	
	Modified area(Δ_3) (γ =0.0783)	0.6064	0.34235	0.25433	0.21032	0.18392	1.59733
	Final weights ($W_{3j(Final)}$)	0.37963	0.21433	0.15922	0.13167	0.11514	1.00
4	Cumulative Proportions	0.04950	0.16831	0.27723	0.58416	1.00	
	Area under N(0,1)	0.5199	0.5675	0.6103	0.7190	0.8413	
	Modified area(Δ_4) (γ =0.507426)	0.5199	0.31352	0.24472	0.21032	0.18969	1.47815
	Final weights ($W_{4j(Final)}$)	0.35172	0.21210	0.16556	0.14229	0.12833	1.00
5	Cumulative Proportions	0.05941	0.18812	0.25743	0.58416	1.00	
	Area under N(0,1)	0.5239	0.5753	0.6026	0.7190	0.8413	
	Modified area(Δ_5) (γ =0.1058)	0.5239	0.31485	0.24517	0.21032	0.18942	1.48366
	Final weights ($W_{5j(Final)}$)	0.35311	0.21221	0.16524	0.14176	0.12767	1.00
6	Cumulative Proportions	0.28713	0.42574	0.54455	0.79208	1.00	

	Area under N(0,1)	0.6141	0.6664	0.7054	0.7852	0.8413	
	Modified area(Δ_6) (γ =0.075733)	0.6141	0.34492	0.25519	0.21032	0.18341	1.60794
	Final weights ($W_{6j(Final)}$)	0.38192	0.21451	0.15871	0.13080	0.11406	1.00
7	Cumulative Proportions	0.08911	0.22772	0.28713	0.60396	1.00	
	Area under N(0,1)	0.5359	0.5910	0.6141	0.7257	0.8413	
	Modified area(Δ_7) (γ =0.1018))	0.5359	0.31885	0.2465	0.21032	0.18862	1.50019
	Final weights ($W_{7j(Final)}$)	0.35722	0.21254	0.16431	0.14020	0.12573	1.00

Calculation of weights to response categories of Items: Method 3

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

Observations:

- Equidistant property is satisfied for Method 2 and 3.
- Inter Quartile range (IQR) ($Q_3 - Q_1$) for Method 2 and Method 3 was 1.28181 and 3.33277 respectively. Number of outliers defined as observations that fall below Q_1− 1.5(IQR) or above Q_3 + 1.5 (IQR) was 3 for Method 2 and Nil for Method 3.

4.2 BREAKING OF TIES

Method 1 resulted in large number of ties of subject scores. Length of tie ranged between 2 to 12 over 14 levels of scores. For Method 2, scores of only two subjects were tied at 3.57363 and no tie was found in Method 3

Consideration of different weights to different response categories of different items by Method 2 and 3 resulted in breaking ties of subject scores in Method 1. Thus, Method 2 an3 distinguish the same summative score on the basis of how the score was obtained. For example, six subjects who scored 20 each in Method 1, obtained different scores for Method 2 and 3 as can be seen from the Table- 3.

Table – 3

Tied Scores of Method 1 and corresponding scores in Method 2 and 3 (Example)

Item-wise scores	**Total score as per**

Item 1	**Item 2**	**Item 3**	**Item 4**	**Item 5**	**Item 6**	**Item 7**	**Method 1**	**Method 2**	**Method 3**
2	5	1	4	1	3	4	20	4.30484	3.284806
4	1	1	4	4	1	5	20	5.092923	3.500008
2	1	1	4	3	4	5	20	4.634009	3.422747
5	5	2	4	1	2	1	20	4.412625	3.304028
2	2	2	4	4	1	5	20	4.779016	3.431687
4	1	2	3	4	1	5	20	4.91835	3.476549

4.3 RELATIONSHIPS OF SCORES BY VARIOUS METHODS:

For the *i*-th item, a score of j as per Method 1 becomes $j.W_{ij(Final)}$ in Method 2 and Method 3. Thus, scores under the three methods are linearly related. However, $W_{ij(Final)}$ are different for different item – response category combinations and different methods. For illustration, relationships of score of Item 1 under the three Methods are shown in Table – 4.

Table – 4

Relationships of scores of Item 1 and the proposed Methods

Scores		
Method 1	**Method 2**	**Method 3**
1	0.023056	0.350358
2	2(0.185888)= 0.371776	2(0.211992) = 0.423984
3	3(0.240165) = 0.720495	3(0.16587) = 0.49761
4	4(0.267304) = 1.069216	4(0.142808) = 0.571232
5	5(0.283587)= 1.417935	5(0.128972)= 0.64486

Clearly, scores of Item 1 by Method 2 and 3 are equidistant, since

$5W_{15(Final)} - 4W_{14(Final)} = 4W_{14(Final)} - 3W_{13(Final)} = 3W_{13(Final)} - 2W_{12(Final)}$

$= 2W_{12(Final)} - W_{11(Final)}$ = 0.34872 for Method 2

= 0.07362 for Method 3

If M_i denote subject scores obtained by the *i*-th Method, for *i* =1, 2, 3, then regression equations on M_1 are as follows**:**

$M_2 =$ (-) 0.7973 + 0.27549 M_1 where $R^2 = 0.9492$

$M_3 =$ 2.343404 + 0.0541 M_1 where $R^2 = 0.7325$

$M_2 =$ (-) 8.57656 + 3.941475M_3 where $R^2 = 0.7763$

High value of R^2 indicate goodness of fit of the data to the linear model. The figure 1 indicates almost linearity of subject scores by the three methods

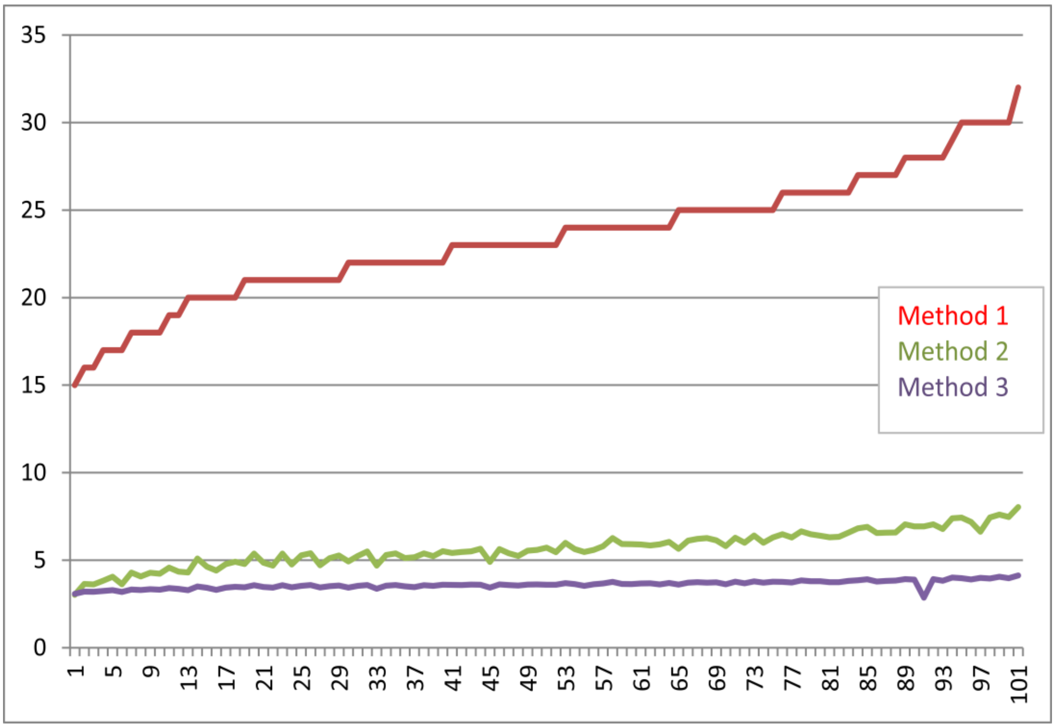

Fig. 1: Scores of subjects by three methods

4.4 DESCRIPTIVE STATISTICS

Descriptive statistics for method 1, method 2 and method 3 are shown in Table – 5.

Table – 5

Descriptive statistics for various Methods

Description	Method 1 (Summative score)	Method 2 (Based on frequency of response category	Method 3 (Based on area under $N(0,1)$)
Test mean	23.40594	5.65081	3.60965
Range of Item mean	[2.06931, 4.15842]	[0.41171,1.12446]	[0.43435, 0.58289]
Test variance	12.38356	0.99015	0.04948
Range of item variance	[1.21465, 2.38752]	[0.08476, 0.18113]	[0.00159,0.01431]
Test skewness	0.003312	-0.09789	-0.34659
Range of Item skewness	[-1.46419, 1.02592]	[-1.46419, 1.02592]	[-5.34448, 1.02592]
Test Kurtosis	1.39444	-0.16945	0.59986

Range of Item Kurtosis	[-1.54227, 1.39444]	[-1.54227, -0.11678]	[-1.54227, 43.96021]

Observations for Method 2 and 3:

- Significant reduction of mean and variance for the test as well as for the items.
- The weighted sum scores made the data more homogenous. Method 3 had minimum variance.
- Test skewness was close to zero for each method. However, same was not found to be true for the item 3 by Method 3. Ratio of sample skewness and Standard Error of Skewness (SES) for the Methods 1, 2 and 3 were 0.013787, -0.40752 and -1.44283 respectively. Thus, the population is very unlikely to be skewed either positively or negatively.
- Test kurtosis (excess of 3) had low values for Methods 2 and 3. However, kurtosis of the summative score method was highest (around 1.40). Ratio of sample kurtosis and Standard Error of Kurtosis (SEK) exceeded 2 for each of the three methods implying that test scores of each method had positive excess kurtosis (leptokurtic).

4.5 ITEM RELIABILITY

A subject score of j in i-th item as per the Method 1, translates into $j.W_{ij(Final-Method\ 2)}$ and $j.W_{ij(Final-Mathod\ 3)}$ respectively $\forall\, j = 1,2,3,4,5$. However, weights vary across the items and response categories. Linearity of the three sets of scores and different weights to response categories of different items is likely to keep correlation between a pair of item unchanged for some of the items but not for all. However, Item reliability in terms of correlation between item scores and test scores differed for different Methods due to combined effect of different weights. Table 6 gives Item-total correlations for the three methods.

Table – 6

Item reliability for various Methods

	Item reliability (Item-total correlation)		
	Method 1	**Method 2**	**Method 3**
Item 1	0.261721	**0.362304**	0.315228
Item 2	**0.298098**	0.213792	0.143324
Item 3	0.396144	0.330685	**0.471869**
Item 4	0.460183	**0.518658**	0.487556
Item 5	0.385952	**0.412299**	0.362374
Item 6	**0.465345**	0.327931	0.263309
Item 7	0.390203	**0.458053**	0.450303

Legend: Max. Correlations (Item reliability) are given in bold

4.6. RANK CORRELATION OF ITEM RELIABILITY

Spearman ρ between Item reliability obtained by various methods are shown in the Table – 7

Table - 7
Rank Correlation Matrix (Spearman ρ) of Item reliability

	Method 1	Method 2	Method 3
Method 1	1.0	0.14286	0.39286
Method 2		1.0	0.78571
Method 3			1.0

Observations: Rank correlation (Spearman ρ) of item reliability was maximum between Method 2 and Method 3.

4.7 TEST OF NORMALITY

Anderson – Darling test for Normality was used to test H_0**:** subject scores follow Normal distribution. Values of test statistic and associated *p*-values for each of the four methods are shown in the Table – 8.

Table - 8
Anderson – Darling test of Normality

	AD statistics	***p*-values**	**Remarks**
Method 1	0.504	0.199148427	I H_0 is accepted
Method 2	0.163	0.942107874	H_0 is accepted
Method 3	0.294	0.593228512	H_0 is accepted

4.8. FACTURE STRUCTURES FOR THE METHODS

PCA and also FA with varimax rotation and Kaiser Normalization resulted in 4 independent factors explaining cumulative variance of around 66% for each of the three methods. The rotated component matrix (loadings) containing estimates of correlations between items and estimated components for the various methods are given in Table – 9.

Table - 9
Rotated Component Matrix for the three methods

	Components			
	1	**2**	**3**	**4**

Item 1	-0..065 (-0.251) [-0.133]	0.060 (0.146) **[0.756]**	**-0.711 (-0.697)** [0.140]	0.114 (-0.132) [-0.172]
Item 2	**0..875 (0.905) [0.812]**	0.181 (0.154) [-0.117]	0.195 (-0.017) [0.251]	-0.126 (0.027) [0.263]
Item 3	**-0.581** (-0.382) [-0.112]	0.349 (0.259) **[-0.643]**	0.461 (0.733) [0.166]	-0.323 (-0.157) [-0.407]
Item 4	0.079 (0.075) [0.041]	**0.935 (0.944)** [-0.011]	-0.086 (0.031) **[0.913]**	0.044 (-0.008) [-0.096]
Item 5	0.019 (0.144) [0.470]	-0.011 (-0.136) [0.343]	0.071 (0.146) [-0.088]	**-0.684 (-0.637)** [-0.205]
Item 6	0.102 (0.119) [-0.032]	-0.102 (-0.150) [-0.035]	**0.584** (0.227) [-0.039]	**0.527 (0.736) [0.865]**
Item 7	-0.366 (-0.450) **[-0.563]**	0.329 (0.330) [0.137]	-0.056 (-0.115) [0.417]	0.459 (0.308) [0.213]

Legend: Substantial loadings are given in bold

Note: Figures without parentheses denote factor loadings by Method 1; figures within ( ) denote factor loadings by Method 2 and figures within [ ] denote factor loadings by Method 3

Observations:

- Load of an item to a component was different for different Methods.
- Item 1 had maximum load with the 2nd component by Method 3 followed by 3rd component in Method 1 and 2
- Item 2 was heavily loaded with the 1st component by all the three methods
- Item 3 had substantial load to 1st component and 2nd component in Method 3.
- Loads of the Item 4 were substantial for 2nd component (Method 1 and 2) and 3rd component(Method 3)
- Item 5 had high load to the 4th component by Method 1 and 2
- Item 6 was heavily loaded with the 4th components by all the three Methods and with 3rd component (Method 1)
- Item 7 showed substantial load with the 1st component by Method 1 only

4.9 CORRELATIONS AMONG THE METHODS

Subject scores obtained by each of the three methods are linear in nature and thus, high correlations are likely between a pair of methods. Correlation matrix of the three methods is given in Table – 10.

Table 10
Correlation matrix of the three Methods

	Method 1	Method 2	Method 3
Method 1	1	0.97427	0.85586
Method 2		1	0.88110
Method 3			1

Observations: Almost linear relationships among the methods resulted in high correlation between each pair of methods.

5. CONCLUSIONS

The proposed methods of scoring Likert scale as weighted sum where different weights are assigned to different response categories using only the frequencies or empirical probabilities of Item – Response categories resulted in continuous data satisfying equidistant property and providing platform to perform analysis in parametric set up. If frequency of a particular response category of an item is zero, the method may fail and can be taken as zero value for scoring Likert items as weighted sum. Proposed approaches made the data homogenous and distinguished the same summative score on the basis of how the score was obtained.

All the three methods passed normality test and produced same number of independent factors though factor loadings were different for different methods.

Each of the proposed method had strong linear relationship with the summative score method (Method 1). Correlations ranged between 0.88 to 0.94. However, Item reliability in terms of correlation between item scores and test scores differed for different Methods due to combined effect of different weights. High correlations among the methods provide a reconciliation of the ordinal - interval controversy, in the sense that there may not be much harm in using data generated from summative scores of Likert questionnaire assuming equal weights which correlates high with the transformed scores as weighted sum satisfying many desired properties.

However, considering the theoretical advantages including meaningfulness of operations and comparisons, platform to undertake parametric statistical analysis and easiness to compute weights, either the Method 2 or 3 may be used for scoring Likert items. Empirically, the Method 3 showed minimum value of variance, no tied scores and no outliers. Thus, Method 3 may be preferred over the Method 2.

After giving weights to response categories, studies may be undertaken to assign further weights to items so as to ensure equal correlation between item score and test score i.e. the test score is equi-correlated with item scores or standardized score of each item, to further justify summation of item scores which are of equal importance and explore relationships of the weighted sum approaches and summative Likert scores when the latter may not follow normal distribution and effect size using simulation studies involving multi data set for generalization.

ACKNOWLEDGMENT

Acknowledgement: Nil

Funding details: Nil

Conflict of interests: Nil

Received at: 18.05.2020, Accepted for publication on: 28.05.2020

REFERENCES:

1. Barua, Ankur (2013). Methods for Decision–making in Survey Questionnaires Based on Likert Scale, *Journal of Asian Scientific Research*, 3(1):35-38

2. Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception.

Psychological Bulletin, 87*,* 564-567.

3. Granberg-Rademacker, J. S. (2010) An Algorithm for Converting Ordinal Scale Measurement Data to Interval/Ratio Scale. *Educational and Psychological Measurement,* 70 (1), 74-90.

4. Hand, D. J. (1996). Statistics and the theory of measurement. *Journal of the Royal Statistical Society, Series A*, 159, 445-492.

5. Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research,* 71, 105-131.

6. Jacoby, W. G. (1999). Levels of measurement and political research: An optimistic view. *American Journal of Political Science,* 43*,* 271-301.

7. Jamieson, S. (2004). Likert scales: How to (ab) use them. *Medical Education, 38,* 1212 -1218

8. King, G, Murray, C. J. L., Salomon, J. A., & Tandon, A. (2003). Enhancing the validity of Cross-cultural comparability of measurement in survey research. *American Political Science Review*, 97, 567-583.

9. Lantz, B. (2013) Equidistance of Likert –Type Scales and Validation of Inferential Methods using Experiments and Simulation, T*he Electronic Journal of Business Research Methods,*11(1), 16-28

10. Narens, L., & Luce, R. D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin,* 99*,* 166-180.

11. Townsend, J. T., & Ashby, F. G. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin*, 96, 394-401.

12. Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *American Statistician*, 47, 65-72.

13. Wilson, T. P. (1971). Critique of ordinal variables. *Social Forces*, 49, 432-444.

14. Wu, Huiping and Leung, Shing-On (2017). Can Likert Scales be treated as Interval Scales? —A Simulation Study, *Journal of Social Service Researc*h, 43:4, 527-532,

15. Zijlmans, Eva A. O., Andries van der Ark, L., Tijmstra, J and Sijtsma, K. (2018) Methods for Estimating Item-Score Reliability, *Applied Psychological Measurement*, 42(7) 553–570

16. Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology*, 34, 390-400.