Journal of Innovation in Psychology, Education and Didactics	Vol. 22, No. 1 2018	23 - 40

RELIABILITY OF LOGISTICS PERFORMANCE INDEX

Satyendra Nath CHAKRABARTTY [a]*

[a] Indian Ports Association, Indian Maritime University

Abstract

A survey using six Likert items was used to obtain six components of Logistics Performance Index (LPI). Each component was rated by respondents on a scale of 1–5. The LPI reports use a single item to measure a complex construct and do not indicate reliability of the entire scale or an item/question. Most of the existing methods for verifying the reliability of Likert scales with different sets of assumptions violate one or more features of such scales. This paper proposes two non-parametric measures based on directional statistics to find reliability of items and the scale used in LPI considering empirical probabilities of Item – Response categories without making any assumptions for the observed variables or the underlying variable being measured. Properties and advantages of the proposed methods are discussed along with empirical verification with a hypothetical data. Use of non-parametric reliability is recommended for Likert-type data for clear theoretical advantages and easiness in calculations.

Key words*:* Angular Association, Bhattacharyya's measure, Likert scale, most preferred direction, polychoric correlations

Introduction

A single index measure of performance of the logistic sector in terms of Logistics Performance Index (LPI) based on survey was developed by the World Bank in 2007 combining scores of six chosen dimensions of trade (or components) to provide a unique reference for better understanding of major trade logistics impediments. LPI has been updated in 2010, 2012, 2014 and 2016. Each version indicates LPI scores and ranking of countries. The number of countries covered under various reports ranges between 150 to 160. LPI score of a country ranges from 1 to

* Corresponding author. Satyendra Nath Chakrabartty, contact number (917) 833-3652
E-mail address: snc12@rediffmail.com

5, where a higher score indicates better performance. The index has gained acceptance among policymakers and professionals at national levels of the world.

Basic data on six components were obtained from on line surveys using six Likert items, each with five response categories. About 1000 respondents were chosen from the logistics professionals, multinational freight forwarders and the main express carriers of 125 countries. The instrument (scale) used in LPI survey is a questionnaire (questions 10–15) to collect the raw data for the international LPI. Thus, the scale can be regarded as a Likert scale consisting of six items to measure six components (3 inputs relating to policy regulations and 3 outputs about service delivery performance), where each component was rated by respondents on a scale of 1–5. The score of 5 implied highest desirable feature of the item or question. Details are presented in Table 1.

Table 1. Components and nature of variables for the international LPI

Question No.	Components	Response Category (1)	Response Category (5)	Nature of variable
10	The efficiency of customs and border clearance	Very low	Very high	Input
11	The quality of trade and transport infrastructure	Very low	Very high	Input
12	The ease of arranging competitively priced shipments	Very difficult	Very easy	Output
13	The competence and quality of logistics services	Very low	Very high	Input
14	The ability to track and trace consignments	Very low	Very high	Output
15	The frequency with which shipments reach consignees within scheduled or expected delivery times	Hardly ever	Nearly always	Output

The relative LPI score was obtained by normalizing the LPI score so that the best performer has the maximum relative LPI score of 100 percent. Details of the survey methodology and index construction methodology can be seen from Arvis, et.al. (2007, 2010, 2012, 2014 and 2016). It may be noted that LPI Reports do not indicate reliability of the entire scale or an item (or question) to measure a complex construct like efficiency of customs and border management clearance, etc. Issues relating to the scale are as follows: use of single item to measure a construct, nature of data and permissible operations, reliability of item and scale, problems associated with finding reliability of Likert items and Scale.

i) Use of single item to measure a construct

The LPI approach used single item to measure a construct which appears to be complex and may include several factors. For example, revenue realization may be more important than minimizing time to issue custom clearance for poor countries. Similarly, developed countries may give more emphasis to check their concern for terrorism which may affect efficiency of customs and border clearance. Multi Criteria Decision making (MCDM) approach used to measure LPI was not considered while measuring the components. However, it is well accepted that Likert scales with summated scores of multi-items are preferred instead of a single item for measuring a construct. One of the four desired characteristics of a summated rating scale proposed by Spector (1992) is that the scale must contain multiple items.

Peter (1981) found that a single item never fully exhausts everything that is meant by a construct. Hair, et.al. (1998) opined that there is a potential risk of misleading results by selecting a single statement to represent a more complex construct. McIver and Carmines (1981) opined that “It is very unlikely that a single item can fully represent a complex theoretical concept or any specific attribute for that matter” and single item measures tend to be less valid, less accurate, and less reliable. Nunnally and Bernstein (1994) observed that measurement error using multi items averages out when individual scores are summed to obtain a total score and an individual item can only categorize people into a relatively small number of groups and thus lacks scope and precision. In other words, a single item cannot discriminate among fine degrees of a construct and has considerable measurement error. Use of scales with multi items is now standard practice e.g., Netemeyer, Bearden, and Sharma (2003). The issues relating to single-item versus multiple-item was investigated by Blalock (1970) and he remarked that with a single item measure of each variable, one is not aware of measurement error and there is no substitute for the use of multiple measures of important variables. Diamantopoulos, et.al (2003)found that that single item measures in most empirical settings is a risky decision since the circumstances that may favor their use are unlikely to be frequently encountered in practice and recommended to follow scales with multiple items in empirical investigations, as suggested by Churchill (1979); DeVellis (2003).

ii) Nature of data and permissible operations

Data generated from each item with five categories as used in LPI are not continuous but ordinal in nature. Whether Likert-type scales should be regarded as ordinal or interval has been a subject of debate (e.g. Carifio and Perla, 2007; Jamieson, 2004; Michell, 1986); the scaling obtained from the Likert procedure is, certainly, at least ordinal. The response categories tend to be sequential, however not linear; the distances between successive pairs of categories are not the same – an essential property of the Interval scale. The categories formed by the 5-response categories do not tend to satisfy the interval scale assumption which means that summation is not meaningful for such category-based scale. Parametric statistical methods like factor analysis, hierarchical linear models, structural equation models, the t-test, ANOVA, etc. rely on the assumption of normally distributed interval-level data which may not apply to data generated by a Likert scale. Hence, analysis of data obtained by means of such scales is limited to frequency tables, with relative and cumulative relative frequencies, proportions or empirical probability associated with each Question-Response category.

According to Lantz (2013), respondents do not generally perceive a Likert-type scale as equidistant. There have been elaborated some methods to ‘rescale’ ordinal scales to get interval properties (e.g. Granberg-Rademacker, 2010; King et al., 2003; Harwell and Gatti, 2001; Bendixen and Sandler, 1995). But, the use of such methods in practical analysis of Likert-type data is scarce. Besides interval properties, normality and homoscedasticity assumptions also need discussion. According to Chien-Ho Wo (2007), translating Likert-scale data into numerical scores through the Snell’s scaling procedure does not contribute much to pass the normality test. The ALSOS method by Jacoby (1999), an inherently model-drive technique assumes that survey respondents construe questions in a similar way and if the model is incorrectly specified, the scaled variables generated from iterative process could be more biased. According to the item response theory, even large ordinal scales can be radically nonlinear. Granberg-Rademacker, (2010) proposed the Monte Carlo Scaling Method based on a multivariate normal distribution. Regarding the Anchoring Vignettes (AV) approach introduced by King et al. (2003) for correcting the differential item functioning problem, it may be necessary to have more than one vignette for a given latent attitude or variable. In addition, vignette responses are not available in secondary datasets. Muraki (1992) observed that if the data fit the Polytomous Rasch Model and fulfill the strict formal axioms of the said model, it may be considered as a basis for obtaining interval level estimates of the continuum.

iii) Reliability of item and scale

LPI Reports do not indicate value of reliability of the items or the entire questionnaire (scale). It is well known that the standards for reporting research include among others the need to report reliability of test scores.

iv) Problems associated with finding reliability of Likert items and Scale

There are problems associated with finding the reliability of Likert scales. The reliability of a scale quantifies the degree of dependability, consistency or stability of the score when such scale is administered. The test-retest approach is a popular method to find reliability of a Likert Scale. Here, reliability is estimated by computing the correlation of summated scores administered to the same set of respondents on two different occasions. Assumptions made for calculation of correlations are at least interval measurement of the variables and data are continuous and normally distributed, which are generally not satisfied by data generated from the Likert scale. One can get different values of such reliability of the same scale depending on the time gap between the two administrations. Moreover, Berchtold (2016) opined that the term test-retest covers two different concepts namely reliability and agreement. While reliability may reflect ability of a measure to produce the same rankings on both occasions, agreement may require the measure to come out with identical values on both occasions. Thus, interpretation of difference between two successive scores could be due to change of the respondents in the time gap or due to the characteristics of the scale.

In Cronbach's alpha terms, reliability implies that the measurement is continuous, with uncorrelated errors and following normal distribution. When assumption of the continuous nature of data and normality is not confirmed, the variance-covariance matrix may be seriously distorted particularly if two variables manifest themselves in skewed distribution of observed responses (e.g., Flora & Curran, 2004): skewed and/or leptokurtic distributions produce negative bias when the coefficient α is calculated (Sheng and Sheng, 2012; Green and Yang, 2009). Cronbach's alpha value can be increased by adding more items, but deletion of items leads to increased alpha value; redundancy of items may explain a too high alpha value (Streiner, 2003). Limitations and misuse of Cronbach's alpha have often been reported (Cortina, 1993; Schmitt, 1996; Sijtsma, 2009; Eisinga, Te Grotenhuis & Pelzer, 2012; Ritter, 2010).

Possible options could be to replace the correlation matrix of the items by polychoric correlations. For binary items, bi-serial or point bi-serial correlations between an item scores and total scores are more appropriate. Similarly, a polychoric correlation may be computed for correlation between a pair of Likert items or between a Likert item and total score for more accurate estimates of the relationship of the underlying variables (Carroll, 1961). Jöreskog and Sörbom (1996) found that polychoric correlations were the most consistent and robust estimator. Holgado, et. al. (2010) demonstrated advantages of polychoric correlations over Pearson correlations when carrying out analysis of data emerging from a Likert scale. But, polychoric correlation assumes that the two items follow bivariate normal distribution which needs to be tested empirically by goodness of fit tests like the likelihood ratio, chi-squared test, G^2 test, etc. Deviation level from bivariate normality can generate biased estimate of polychoric correlations. Polychoric correlations performed worst on all goodness–of–fit criteria (Babakus, Ferguson and Joreskog, 1987). The distribution of underlying variables can be highly skewed, which may introduce bias in the result of chi-square test to assess goodness of fit of structural equation models (Muthen, 1993). If a polychoric correlation matrix is not definitely positive, problems may occur. Polychoric correlation offers rather unstable estimates for small samples whereas for large samples, the estimates are noisy if there are few empty cells. For items with smaller number of response categories, polychoric correlation between latent continuous variables tends to be attenuated.

Ordinal reliability is referred to as nonparametric reliability coefficients in a nonlinear classical test theory sense even though such reliabilities assume that the underlying variable is continuous (Lewis, 2007). A measure of reliability regarding the coefficient theta (Armor, 1974), based on principal components analysis (PCA) was proposed by Zumbo, Gadermann and Zeisser (2007). If the single factor solution is reasonable for the items, then $\theta = \frac{p}{p-1}\left(1 - \frac{1}{\lambda_1}\right)$where λ_1 is the largest eigen value obtained from the PCA of the correlation matrix for the items. However, estimation of λ_1 based on the sample covariance matrix is extremely sensitive to outlying observations. Gadderman, Guhn and Zumbo (2012) proposed ordinal alpha for ordinal data based on the polychoric correlations and defined ordinal alpha as $\alpha = \frac{p}{p-1}\left(1 - \frac{p}{p + \Sigma\Sigma r_{ij}}\right)$ where p denotes the number of items and r_{ij} denotes the polychoric correlation between items i and j. However, strictly speaking, reliability using polychoric correlation is not a 100% non-parametric approach because of the assumption of bivariate normality of the underlying variables.

Each of the above method has certain advantages and disadvantages. However, the estimation of reliability of a Likert scale under each such method with different sets of assumptions deviates differently and thus gives different values for a single Likert scale. This motivates a need to find methods of obtaining reliability of a Likert item and Likert scale from a single administration of the questionnaire in a non-parametric approach without involving assumptions on the nature or distribution of observed or underlying variables.

Objective

To obtain non-parametric measures of reliability of Likert items and Likert scales as used in LPI to obtain basic data for six identified components from a single administration, using only the permissible operations for a Likert scale, i.e. considering the cell frequencies or empirical probabilities of Item – Response categories without making any assumptions of continuous nature or linearity or normality for the observed variables or the underlying variable being measured along with discussion of properties of such measures and comparison.

Formal description

Suppose there are *n* – respondents who answered each of the *m*-items of a Likert questionnaire where each item has *k*-numbers of response categories. For instance, the case of LPI, *m=6* and *k=5*.

Consider the item-response category frequency matrix $\left(\left(f_{ij}\right)\right)$ of order 6X5 where f_{ij} denotes frequency of the *j*-th response category of the *i*-th item, *i=1,2,…6* and *j= 1,2,…5.* Row total, i.e. $\sum_{j=1}^{5} f_{ij} = n \quad \forall\, i = 1,2,\ldots 6$ and column total ,i.e. $\sum_{i=1}^{6} f_{ij} = f_{0j}$ gives total frequency of the *j*-th response category$\forall\, j = 1,2,\ldots 5$. In other words, f_{0j} denotes the total number of times the *j*-th response category was chosen by all the respondents. Clearly $\sum_{i=1}^{6}\sum_{j=1}^{5} f_{ij} = 6n$= (Sample size). (Number of items)

After administration of the questionnaire to a large number of respondents, one can calculate the 5-dimensional vector of empirical probabilities for the *i*-th item with 5- response categories as $\boldsymbol{P}_i = (\,p_{i1}, p_{i2}, \ldots\ldots, p_{i5})^T$ where p_{ij} is the empirical probability of the *i*-th item and *j*-th the response category and is equal to $\frac{f_{ij}}{n}$. Clearly, $\sum_{j=1}^{5} p_{ij} = 1$. In other words, the vector $\boldsymbol{P}_i$ corresponding to the *i*-th item can be found as $\boldsymbol{P}_i = (\frac{f_{i1}}{n}, \frac{f_{i2}}{n}, \frac{f_{i3}}{n}, \frac{f_{i4}}{n}, \frac{f_{i5}}{n})^T$

Similarly, for the entire questionnaire, the vector showing empirical probabilities of the response categories can be computed as

$\boldsymbol{T} = (\frac{f_{01}}{6n}, \frac{f_{02}}{6n}, \ldots\ldots, \frac{f_{05}}{6n})^T = (\frac{\sum_{i=1}^{6} f_{i1}}{6n}, \frac{\sum_{i=1}^{6} f_{i2}}{6n}, \frac{\sum_{i=1}^{6} f_{i3}}{6n}, \frac{\sum_{i=1}^{6} f_{i4}}{6n}, \frac{\sum_{i=1}^{6} f_{i5}}{6n})^T$. Clearly, $\sum_{j=1}^{5} \frac{f_{0j}}{6n} = 1$

Computation of the vectors P_i's and the vector T form the starting point for further calculation to obtain reliability of the scale and the item reliabilities.

Methodology for the proposed methods

Reliability by Angular association

Association between the *i*-th and *j*-th item can be found by evaluating $Cos\theta_{ij}$ where θ_{ij} is the angle between the vectors $\boldsymbol{P_i}$ and $\boldsymbol{P_j}$ by the formula

$$Cos\theta_{ij} = \frac{P_i P_j^T}{\|P_i\|\|P_j\|} \quad \ldots\ldots \quad \ldots\ldots \quad (1.1)$$

Similarly, item reliability in terms of item-test correlation between the *i*-th item and total score can be obtained by $Cos\theta_{iT}$ where θ_{iT} is the angle between the vectors $\boldsymbol{P_i}$ and $\boldsymbol{T}$

$$Cos\theta_{iT} = \frac{P_i T^T}{\|P_i\|\|T\|} \quad \ldots\ldots \quad \ldots\ldots \quad (1.2)$$

Note that $Cos\theta_{ij}$ as defined in (1.1) satisfy the following:

- If $\boldsymbol{P_i} = \boldsymbol{P_j}$ for the *i*-th and *j*-th item where $i \neq j$ then $Cos\theta_{ij} = 1$ and vice versa.
- $Cos\theta_{ij} = 0$ if and only if the vectors $\boldsymbol{P_i}$ and $\boldsymbol{P_j}$ are orthogonal
- Symmetric i.e. $Cos\theta_{ij} = Cos\ \theta_{ji}$
- Satisfy non-negativity condition i.e. $Cos\theta_{ij} \geq 0$
- Does not satisfy triangle inequality i.e. it does not satisfy $Cos\theta_{XY} + Cos\theta_{YZ} \geq Cos\theta_{XZ}$ where $X \neq Y \neq Z$. In other words, $Cos\theta_{ij}$ is not a metric.

Test reliability should not be computed as average of $Cos\theta_{ij}$'s or $Cos\theta_{iT}$'s since $Cos\theta_{ij}$ does not obey triangle inequality and hence are not additive. The symmetric matrix showing $Cos\theta_{ij}$'s may be used to find value of test reliability and to undertake factor analysis. However, $Cos\theta_{iT}$ will indicate reliability of the *i*-th item.

Following Gadderman, Guhn and Zumbo (2012), reliability of the test can be found by replacing the polychoric correlation between items i and j by $Cos\theta_{ij}$ in the following equation

$$r_{tt} = \frac{m}{m-1}\left(1 - \frac{m}{m + \sum\sum_{i \neq j} Cos\theta_{ij}}\right) \quad \ldots\ldots \quad \ldots\ldots \quad (1.3)$$

where *m* denotes the number of items

For instance, the case of LPI, *m=6* and the formula for reliability of the LPI Scale is

$$r_{tt} = \frac{6}{5}\left(1 - \frac{6}{6+\sum\sum_{i\neq j} Cos\theta_{ij}}\right) \quad \ldots\ldots \quad \ldots\ldots \quad (1.4)$$

Clearly, equation (1.3) and (1.4) requires computation of inter-item correlation matrix in terms of $Cos\theta_{ij}$

It may be noted that reliability as per equation (1.3) does not help to find reliability of the test as a function of item reliabilities.

To make $\boldsymbol{P}_i$ and $\boldsymbol{P}_j$ as unit vector, one may choose $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}_j$ where $\boldsymbol{\pi}_i = \sqrt{\frac{\boldsymbol{P}_i}{\|P_i\|}}$ and $\boldsymbol{\pi}_j = \sqrt{\frac{\boldsymbol{P}_j}{\|P_j\|}}$ so that $\|\pi_i\|^2 = \|\pi_j\|^2 = 1$. In that case cosine of the angle between $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}_0$ becomes the Bhattacharyya's measure.

Bhattacharyya's measure

Association between *i*-th item and *j*-th item with vector $\boldsymbol{P}_i = (p_{i1}, p_{i2}, p_{i3}, p_{i4}, p_{i5})$ and vector $\boldsymbol{P}_j = (p_{j1}, p_{j2}, p_{j3}, p_{j4}, p_{j5})$ can be found by Bhattacharyya's measure (Bhattacharyya, 1943) as:

$$\rho\left(\boldsymbol{\pi}_i . \boldsymbol{\pi}_j\right) = Cos\,\emptyset_{ij} = \sum_{s=1}^{5} \sqrt{\pi_{is}\pi_{js}} \quad \ldots\ldots \quad (1.5)$$

where $\pi_{is} = \sqrt{\frac{p_{is}}{\|P_i\|}}\ \forall\ i = 1,2,\ldots,6 \text{ and } s = 1,2,\ldots.5$

The Bhattacharyya's measure is in fact cosine of the angle $\emptyset_{ij}$ where $\emptyset_{ij}$ is the angle between the two vectors $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}_j$ since $\|\boldsymbol{\pi}_i\|^2 = \|\boldsymbol{\pi}_j\|^2 = 1,$. Thus, it is a measure of similarity between $\boldsymbol{P}_i$ and $\boldsymbol{P}_j$.

Item reliability will be Item-test correlation using Bhattacharyya's measure computed from

$$\rho(\boldsymbol{\pi}_i . \sqrt{\boldsymbol{T}}) = \sum_{j=1}^{5} \sqrt{\frac{f_{Ij}}{mn} p_{Ij}} \quad \ldots \quad \ldots. \quad (1.6)$$

It can be proven easily that

i) the measure is defined even if a p_{ij} is equal to zero i.e. if all respondents do not choose a response category of an item.

ii) if the vectors $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}_j$ are identical, then $\rho\left(\boldsymbol{\pi}_i.\boldsymbol{\pi}_j\right) = 1$. If $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}_j$ are orthogonal, then $\rho\left(\boldsymbol{\pi}_i.\boldsymbol{\pi}_j\right) = 0$.

iii) $0 \leq \rho\left(\boldsymbol{\pi}_i.\boldsymbol{\pi}_0\right) \leq 1$ using Jensen's inequality (Cover and Thomas, 1991).

iv) does not satisfy triangle inequality (Fukunaga, 1990).

Hence, test reliability should not be computed as average of $\rho\left(\boldsymbol{\pi}_i.\boldsymbol{\pi}_j\right)$ since the measure is not a metric. While dealing with vectors of unit length, Rao (1973) has shown that mean and dispersion of the angles $\emptyset_1, \emptyset_2, \emptyset_3, \dots \dots . . \emptyset_k$ can be found as follows:

Mean or most preferred direction is estimated by $\overline{\emptyset} = Cot^{-1}\frac{\sum \cos \emptyset_i}{\sum \sin \emptyset_i}$ and the dispersion by $\sqrt{1 - r^2}$ where $r^2 = \left(\frac{\sum Cos\emptyset_i}{k}\right)^2 + \left(\frac{\sum Sin\emptyset_i}{k}\right)^2$

Reliability of the Likert scale can be defined as $Cos(\overline{\emptyset}) = Cos\left(Cot^{-1}\frac{\sum \cos \emptyset_i}{\sum \sin \emptyset_i}\right)$…. …. (1.7)

The above will help to find

- Reliability of the Likert scale as a function of item reliabilities
- Range of reliability of the Likert scale can be found from

$(Cos\overline{\emptyset} \pm C\sqrt{1 - r^2})$ where *C* is a suitably chosen constant ….. ….. (1.8)

Empirical verification

Raw data in terms of responses received from the respondents on the six items of LPI is not readily available. LPI reports give relative LPI score after normalization. Hence, empirical verification of the two proposed methods is undertaken with a hypothetical data involving 1000 respondents on six items, each with five response categories. Here, the number of items m = 6, k=5 and sample size n = 1000 (Table 2).

Table 2. Item – Response Categories frequency matrix and Probabilities

Items	Frequency/ Probability	RC-1	RC- 2	RC- 3	RC- 4	RC- 5	Total
1	Frequency	190	320	350	110	30	1000
	Probability(P_{1j})	*0.19*	*0.32*	*0.35*	*0.11*	*0.03*	*1.00*
2	Frequency	70	330	340	190	70	1000
	Probability(P_{2j})	*0.07*	*0.33*	*0.34*	*0.19*	*0.07*	*1.00*
3	Frequency	340	110	50	140	360	1000
	Probability(P_{3j})	*0.34*	*0.11*	*0.05*	*0.14*	*0.36*	*1.00*
4	Frequency	100	140	380	300	80	1000
	Probability(P_{4j})	*0.10*	*0.14*	*0.38*	*0.30*	*0.08*	*1.00*
5	Frequency	40	310	370	200	80	100
	Probability(P_{5j})	*0.04*	*0.31*	*0.37*	*0.20*	*0.08*	*1.00*
6	Frequency	*80*	*250*	*290*	*300*	*80*	*1000*
	Probability(P_{6j})	0.08	0.25	0.29	0.30	0.08	1.00
Total	Frequency(f_{0i})	**820**	**1460**	**1780**	**1240**	**700**	**6000**
	Probability($\frac{f_{0i}}{mn}$) =Probability(t_j)	***0.1367***	***0.2433***	***0.2967***	***0.2067***	***0.1167***	***1.00***

Legend: RC- j denotes j-th Response Category

For Angular association

Vector $\boldsymbol{P_1} = (0.19, 0.32, 0.35, 0.11, 0.03)^T$

$$\boldsymbol{P_2} = (0.07, 0.33, 0.34, 0.19, 0.07)^T$$

$$\boldsymbol{P_3} = (0.34, 0.11, 0.05, 0.14, 0.36)^T$$

$$\boldsymbol{P_4} = (0.10, 0.14, 0.38, 0.30, 0.08)^T$$

$$\boldsymbol{P_5} = (0.04, 0.31, 0.37, 0.20, 0.08)^T$$

$$\boldsymbol{P_6} = (0.08, 0.25, 0.29, 0.30, 0.08)^T$$

$$\boldsymbol{T} = (0.1367, 0.2433, 0.2967, 0.2067, 0.1167)^T$$

Here, the length of the vectors are $\|P_1\| = 0.5235$, $\|P_2\| = 0.5200$, $\|P_3\| = 0.5286$, $\|P_4\| = 0.5200$, $\|P_5\| = 0.5301$, $\|P_6\| = 0.4956$ and $\|T\| = 0.471441$

Inter-item correlations in terms of $Cos\theta_{ij} = \frac{P_i^T P_j}{\|P_i\|\|P_j\|}$ and item-total correlations in terms of $Cos\theta_{iT} = \frac{P_i^T . T}{\|P_i\|\|T\|}$ are shown below (Table 3).

Table 3. **Inter – item & Item – total correlations in terms of** $\mathbf{Cos\theta_{ij} \& Cos\theta_{iT}}$

	Item-1	Item-2	Item-3	Item-4	Item-5	Item-6	Test ($Cos\theta_{it}$)
Item-1	1.00	0.9585	0.5186	0.8531	0.9395	0.8888	0.9479
Item-2		1.00	0.4689	0.9061	0.9958	0.6229	0.9716
Item-3			1.00	0.5064	0.4390	0.5252	0.6382
Item-4				1.00	0.9229	0.9607	0.9456
Item-5					1.00	0.9637	0.9657
Item-6						1.00	0.9733
Test							1.00

Observations:

- Item correlation in terms of $Cos\theta_{ij}$ was positive ∀ *i*= 1,2,….,6 and *j*=1,2,……..,5
- Each item reliability in terms of $Cos\theta_{iT}$ was positive and close to unity, except for the 3rd Item.

Here, the sum of inter-item correlations excluding the diagonal elements,

$\sum\sum Cos\theta_{ij}$ for $i \neq j$ is 2(11.4701) = 22.9402

From (1.3), reliability of the test $r_{tt} = \frac{m}{m-1}\left(1 - \frac{m}{m+\sum\sum_{i\neq j} Cos\theta_{ij}}\right)$

$= \frac{6}{5}\left(1 - \frac{6}{6+24.9633}\right) = 0.9512 \approx 0.95$

Item reliabilities computed from Angular Association method are shown in Table 4.

Table 4. Item Reliability as per Angular Separation Method

Item	Item reliability ($Cos\theta_{it}$)	Ranks of items w.r.t. Item reliability
1	0.9479	4th
2	0.9716	2nd
3	0.6382	6th
4	0.9456	5th
5	0.9657	3rd
6	0.9733	1st

For Bhattacharyya's measure

As per (1.5), $\rho\left(\pi_i.\pi_j\right) = Cos\ \emptyset_{ij} = \sum_{s=1}^{5}\sqrt{\pi_{is}\pi_{js}}$ is the correlation between *i*-th and *j*-th item and $Cos\emptyset_{iT}$ is the reliability of the *i*-th item. Values are shown in the Table 5.

Table 5. Inter-Item and Item-Total Correlations as per Bhattacharyya's Measure

	Item-1	Item-2	Item-3	Item-4	Item-5	Item-6	Test ($Cos\emptyset_{iT}$)
Item-1	1.00	0.9585	0.5186	0.8531	0.9395	0.8888	0.9479
Item-2		1.00	0.4690	0.9061	0.9958	0.9623	0.9716
Item-3			1.00	0.5064	0.4390	0.5252	0.6382
Item-4				1.00	0.9229	0.9607	0.9456
Item-5					1.00	0.9637	0.9657
Item-6						1.00	0.9733
Test							1.00

Item reliabilities computed from Bhattacharyya's measure are shown in Table 6.

Table 6. Item reliability as per Bhattacharyya's measure

Item	Item reliability $[\rho(\pi_i.\sqrt{T}]$= $(Cos\emptyset_{iT})$	$(Sin\emptyset_{iT})$	Ranks of items w.r.t $Cos\emptyset_{iT}$
1	0.9479	0.1015	4th
2	0.9716	0.0560	2nd
3	0.6382	0.5927	6th
4	0.9456	0.1058	5th
5	0.9657	0.0674	3rd
6	0.9733	0.0527	1st
Sum	$\sum_{i=1}^{5} Cos\emptyset_{iT}$ =5.4423	$\sum_{i=1}^{5} Sin\emptyset_{iT}$ = 0.9761	

Here, $\overline{\emptyset} = Cot^{-1}\frac{\sum\cos\emptyset_i}{\sum\sin\emptyset_i} = Cot^{-1}\frac{5.4423}{0.9761} = Cot^{-1}\,5.5755$ = 10.168 degree

Thus, reliability of the test is $Cos(\overline{\emptyset}) = Cos\,(Cot^{-1}\frac{\sum\cos\emptyset_i}{\sum\sin\emptyset_i}) =$

$Cos(10.168) = 0.9843 \approx 0.98$

Conclusions

The reliability of a Likert item and Likert scale (as used in LPI) was found from a single administration using only the permissible operations for a Likert scale considering the frequencies or probabilities of Item – Response categories without involving assumptions of the continuous nature or linearity or normality for the observed variables or the underlying variable being measured. Such reliabilities are in fact non-parametric reliabilities. Such non-parametric reliabilities are critically relevant to practitioners and researchers in the social sciences in general and logistics performance studies in particular. Use of non-parametric reliability is recommended for Likert-type data for clear theoretical advantages and easiness in calculations.

Two methods have been presented to find reliability of a Likert item and a Likert scale avoiding the use of polychoric strategy which is akin to a data transformation and quantifying the reliability of the item response data in transformed metric. However, none of the method is a metric since the triangle inequality condition is not satisfied by any of the proposed method. Accordingly, reliability of the test should not be found as average of inter-item correlations or item-total correlations or item reliabilities.

Reliability of an item was considered as association of the vector showing empirical probabilities of response-categories of the item (P_i) and the vector showing empirical probabilities of response-categories of total score (T).

Reliability of the test for the Angular Association method replaced polychoric correlation between items i and j by $Cos\theta_{ij}$ which can be computed irrespective of the nature of distributions of the observed or underlying variables or factor structure. The value of test reliability by the Angular Association method was found to be 0.95.

Test reliability by Bhattacharyya's measure is cosine of the most preferred direction $(\overline{\emptyset})$ among the vectors of unit length which is a measure of average of the angles between π_i and T. Reliability of the test by Bhattacharyya's measure was found to be 0.98.

Significant values of the elements of the inter-item correlation matrix tend to indicate inter-item consistency leading to possible uni-dimensionality which may be confirmed through factor analysis. Ranks of Items with respect to values of Item reliabilities were found to be same for the two proposed methods. Test reliability by Bhattacharyya's measure has a special property as it can be expressed as a function of item reliabilities. The approach also helps to find the range of reliability of the Likert scale as $(Cos\overline{\emptyset} \pm C\sqrt{1-r^2})$ where C is a suitably chosen constant.

Future studies may be undertaken to find item reliability and reliability of the scale used in LPI using raw data (and not relative LPI score after normalization) and to facilitate comparison of the two proposed methods.

References

Arvis, J. F., Mustra, M. A., Panzer, J., Ojala, L., & Naula, T. (2007, 2010, 2012). *Connecting to Compete: Trade Logistics in the Global Economy*. World Bank, Washington DC.

Arvis, J.F., Saslavsky, D., Ojala, L., Shepherd, B., Busch, C., & Raj, A. (2014). *Connecting to Compete: Trade Logistics in the Global Economy*. World Bank, Washington DC.

Arvis, J.F., Saslavsky, D., Ojala, L., Shepherd, B., Busch, C., Raj, A., & Naula, T. (2016). *Connecting to Compete: Trade Logistics in the Global Economy.* World Bank, Washington DC.

Armor, D. J. (1974). Theta reliability and factor scaling. In H. Costner (Ed.), *Sociological methodology* (pp. 17-50). San Francisco: Jossey-Bass.

Babakus, E., Ferguson, C. E., & Joreskog, K. G. (1987). The Sensitivity of Confirmatory Maximum Likelihood Factor Analysis to Violations of Measurement Scale and Distributional Assumptions. *Journal of Marketing Research,* Vol. XXIV, No. 2, pp. 222-228.

Bendixen, M. T., & Sandler, M. (1995). Converting verbal scales to interval scales using correspondence analysis. *Management Dynamics: Contemporary Research*, 4 (1), 31-49.

Berchtold, A. (2016). Test–retest: Agreement or reliability? *Methodological Innovations,* Vol. 9, 1–7.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society*, 35, 99–110.

Blalock, H. M., Jr. (1970). Estimating measurement error using multiple indicators and several points in time. *American Sociological Review*, *35*(1), 101-111.

Carroll, J. B. (1961). The nature of data, or how to choose a correlation coefficient. *Psychometrika*, *26*, 347-372.

Carifio, J. & Perla, R. J. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3, 106-116.

Chien-Ho Wu (2007). An Empirical Study on the Transformation of Likert-scale Data to Numerical Scores. *Applied Mathematical Sciences*, Vol. 1, no. 58, 2851 – 2862.

Churchill, G. A., Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16 (1), 64–73.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology,* 78, 98-104.

Cover, T., & Thomas, J. A. (1991). Elements of Information Theory. *Wiley-Interscience*

Diamantopoulosa, A., Schlegelmilchb, B. B., Sinkovicsd, R.R., & Bohlenc, G. M. (2003). Can socio-demographics still play a role in profiling green consumers? A review of the evidence and an empirical investigation. *Journal of Business Research, 56*, 465 – 480, doi:10.1016/S0148-2963(01)00241-7.

DeVellis, Robert F. (2003). Scale Development: Theory and Applications, 2nd Edition, *Applied Social Research Methods Series*, Vol. 26, SAGE publication.

Eisinga, R., Te Grotenhuis, M., & Pelzer, B. (2012). The reliability of a two-item scale: Pearson, Cronbach or Spearman-Brown?" *International Journal of Public Health*, 58(4), 637-42, doi: 10.1007/s00038-012-0416-3.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9, 466-491.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition* (Second Edition). San Diego, CA, USA: Academic Press Professional, Inc.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research and Evaluation, 17*(3), 1-13.

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.

Green, S. B., & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: an alternative to coefficient Alpha. *Psychometrika* 74, 155–167. doi: 10.1007/s11336-008-9099-3.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis*, fifth edition. New Jersey: Prentice-Hall International, Inc.

Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, *71*(1), 105-131.

Holgado–Tello, F.P., Chacón–Moscoso, S., & Barbero–García, I. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables, *Qual Quant*, *44*, 153–166, DOI 10.1007/s11135-008-9190-y.

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

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

Jöreskog, K.G., & Sörbom, D. (1996). *LISREL 8 User's reference* guide. Chicago: Scientific Software.

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.

Lantz, B. (2013). Equidistance of Likert-Type Scales and Validation of Inferential Methods Using Experiments and Simulations. *The Electronic Journal of Business Research Methods*, 11(1), 16-28.

Lewis, C. (2007). Classical test theory. In C. R. Rao and S. Sinharay (Eds.), *Handbook of Statistics, Vol. 26: Psychometrics,* (pp. 29-43). Amsterdam, The Netherlands: Elsevier Science B.V.

Michell, J. (1986).Measurement Scales and Statistics: A Clash of Paradigms. P*sychological Bulletin*, 100 (3), 398-407.

McIver, J. P., & Carmines, E. G. (1981). *Unidimensional scaling*. Thousand Oaks, CA: Sage.

Muraki, E. (1992). A generalizes partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16,* 159-176.

Muthen, B. O. (1993). Goodness of Fit with categorical and other non-normal variables. In Bollen & Long (Eds.) *Testing Structural Equation Models* (pp. 205-234). Newbury Park: Sage.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks: Sage Publications, Inc.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw- Hill.

Peter, J.-P. (1981). Construct validity: a review of basic issues and marketing practices. *Journal of Marketing Research,* 18(2), 133– 145.

Rao, C. R. (1973) *Linear Statistical Inference and its Application.* 2nd Edition. New Delhi: Wiley Eastern Private Limited.

Ritter, N. (2010). Understanding a widely misunderstood statistic: Cronbach's alpha. *Paper presented at Southwestern Educational Research Association (SERA) Conference*, New Orleans, LA (ED526237).

Spector, P. (1992). *Summated rating scale construction*. Thousand Oaks, CA: Sage.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8,* 350–353, doi: 10.1037/1040- 3590.8.4.350.

Sheng, Y. and Sheng, Z. (2012). Is coefficient alpha robust to non-normal data?. *Frontiers in Psychology,* 3(34), 1-13, doi 10.3389/fpstg.2012.00034.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74,* 107- 120, doi: 10.1007/s11336-008-9101-0.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment,* 80(1), 99-103.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6,* 21-29.