Angular similarity in test parameters

Satyendra Nath Chakrabartty

Abstract
Through N-dimensional person space, the article gives measures of test parameters and item statistics, including difficulty/discriminating value of test, correlations between a pair of items, and item-total correlations with binary items using angular similarity between two vectors. Relationships between difficulty value and discriminating value of items and test were derived, including relationship between test reliability and test discriminating value. Reliability of a test as per theoretical definition in terms of length of score vectors of two parallel subtests and angle between such vectors was derived. The method was extended to find reliability of a battery of tests. Reliability and discriminating value of a Likert-type item and scale was found in terms of angular similarity without involving assumptions of continuous nature or linearity or normality for the observed variables, or the underlying variable being measured. The proposed methods also avoid test of unidimensionality or assumption of normality or bivariate normality associated with the polychoric correlations. Thus, the proposed methods are in fact nonparametric and considered as improvement over the existing ones. Reliability as a measure of association of two vectors and discrimination as a measure of distance between the vectors are likely to show a negative relationship.

Keywords
N-dimensional person space, item statistics, reliability, difficulty and discriminating values, test battery

Introduction
It is common to consider the variables as orthogonal axes and performance of individuals as points in the Euclidean space defined by the axes. However, it is possible to treat each person as an axis and represent the variables as points or vectors in the Euclidean space. This type of presentation is called “subject space” or “person space” or “N-dimensional person space” for N-persons. Such presentation has been considered by Reyment and Joreskog (1993), Huberty (1994), Chong et al. (2002), and so on. The orthogonality of the axes is an accurate indication if the persons are assumed to be independent of one another. Chong et al. (2002) observed that vectors representing latent factors rotate in subject space rather than variable space. In regression analysis, eigenvalues can be conceptualized as subject space equivalent to variance inflation factor (VIF). The concept of subject space can also be used as a means of visualization of spatial relationships. Glass and Collins (1970) dealt with three variables in N-dimensional subject space to find possible values of $r_{XY}$ when $r_{XZ}$ and $r_{YZ}$ are fixed.

If N-persons take a test, then the observed test scores can be viewed as a point or a vector in N-dimensional space. Once a person has taken a test, he or she is not considered to take the test again. Thus, a person axis is not overlapped. The angle $\theta$ between the two N-dimensional vectors $X$ and $Y$ is given by $\cos \theta = \frac{X \cdot Y}{\|X\| \|Y\|}$, where $\|X\|$ denotes length of the vector and is defined as $\|X\| = \sqrt{\sum_{i=1}^{N} X_i^2}$. $\|Y\|$ is defined accordingly. For N-persons, the number of $\cos \theta_{ij}$ for $i \neq j$ is the same as the number of possible pairs, which is $N(N-2)/2$. This gives the novel area of angular statistics, where $\cos \theta$ gives similarity between two vectors of the same dimension. In many diverse scientific fields, measurements are directions. Use of cosine similarity is common in areas like machine learning, classification, data mining, information retrieval, and pattern recognition.

This type of presentation helps to make useful inferences about the sample as well as computation of various parameters of the test, along with geometrical interpretation of such computations. It is possible to give theoretical formulation of...
mental testing through $N$-dimensional person space primarily in terms of the length of the observed score vector, angle separating the vector with the vector representing the maximum possible score in the test. For item statistics, the corresponding two vectors will be item score vector and the vector for maximum possible item score with each component equal to unity for a binary item.

For a test consisting of binary items (1 for right answer and 0 for others), the article attempts to provide geometric visualization along with computation of parameters of the test and item statistics, including test error variance and thereby the test reliability, as a ratio of true score variance and observed score variance, via a single administration of the test under classical test theory. Thus, the approach is an improvement as it enables calculation of theoretically defined reliability despite the fact that true scores of individuals taking the test are not known. Subsequent to the determination of the error variance of the test, test reliability method is extended to find reliability of a battery of tests. The approach can be extended for Likert-type test.

**Methodology**

The setup for test consists of binary items.

Suppose a test consisting of $n$-binary items has been administered among $N$-subjects. Scores obtained by the subjects can be presented as a point or vector $X$ in the $N$-dimensional space with components $X_1, X_2, \ldots, X_N$, where $X_i$ denotes test score of the $i$th subject. One may consider another vector $I$ representing maximum possible score in the test with components $I_1, I_2, \ldots, I_N$, where $I_i = I_j = n$ for $i, j = 1, 2, \ldots, N$. Call the vector $I$ as maximum possible vector or ideal vector. Let the angle between the vectors $X$ and $I$ be $\theta_X$. Similarly, the true score vector $T$ with components $T_1, T_2, \ldots, T_N$ can be conceptualized, where the angle between the vectors $T$ and $I$ is $\theta_T$.

### Dot product

Dot product of two $N$-dimensional vectors $X$ and $Y$ is defined as

$$X \cdot Y = \|X\| \|Y\| \cos \theta$$

(1)

where $\|X\|$ denotes length of the vector $X$ and is defined as

$$\|X\| = \left( \sum_{i=1}^{N} X_i^2 \right)^{1/2}$$

$\|Y\|$ is defined accordingly and $\theta$ is the angle between the vectors $X$ and $Y$.

Note that

1. $X \cdot Y = 0 \iff \theta = 90^\circ \iff X$ and $Y$ are orthogonal (perpendicular)
2. $X \cdot Y = \cos \theta$ if $X = Y = 1$

### Ranks of individuals

Arranging the components of the vector $X$ in decreasing order will give ranks of the individuals who took the test.

### Mean

Let $\theta_X$ be the angle between the score vector and the ideal vector. Then

$$\cos \theta_X = \sum \frac{X_i I_i}{\|X\| \|I\|}$$

or

$$\cos \theta_X = \sum \frac{X_i}{\|X\| \sqrt{N}}$$

(2)

Thus, test mean is equal to the product of length of the score vector and cosine of the angle between the score vector and the ideal vector divided by square root of sample size. From equation (2)

$$\bar{X} = \left( \sum \frac{X_i}{\|X\| \sqrt{N}} \right)^2$$

(3)

Similarly

$$\bar{T} = \left( \sum \frac{T_i}{\|T\| \sqrt{N}} \right)^2$$

(4)

and

$$\bar{T}^2 = \left( \sum \frac{T_i}{\|T\| \sqrt{N}} \right)^2$$

(5)

Since, $\bar{X} = \bar{T}$, equating (2) and (4), we get

$$\sum \frac{X_i}{\|X\| \sqrt{N}} = \sum \frac{T_i}{\|T\| \sqrt{N}}$$

(6)

Equation (6) helps to find $\|T\| \cos^2 \theta_T$ from the data, even if $\|T\|$ or $\cos^2 \theta_T$ are not known.

Equating (3) and (5), we get

$$\sum \frac{X_i}{\|X\| \sqrt{N}} = \cos^2 \theta_X \sec^2 \theta_T$$

(7)

which gives relationship between $\|X\|$ and $\|T\|$. **Ranks of individuals**

Arranging the components of the vector $X$ in decreasing order will give ranks of the individuals who took the test.

### Mean

Let $\theta_X$ be the angle between the score vector and the ideal vector. Then

$$\cos \theta_X = \sum \frac{X_i I_i}{\|X\| \|I\|}$$

or

$$\cos \theta_X = \sum \frac{X_i}{\|X\| \sqrt{N}}$$

(2)

Thus, test mean is equal to the product of length of the score vector and cosine of the angle between the score vector and the ideal vector divided by square root of sample size. From equation (2)

$$\bar{X} = \left( \sum \frac{X_i}{\|X\| \sqrt{N}} \right)^2$$

(3)

Similarly

$$\bar{T} = \left( \sum \frac{T_i}{\|T\| \sqrt{N}} \right)^2$$

(4)

and

$$\bar{T}^2 = \left( \sum \frac{T_i}{\|T\| \sqrt{N}} \right)^2$$

(5)

Since, $\bar{X} = \bar{T}$, equating (2) and (4), we get

$$\sum \frac{X_i}{\|X\| \sqrt{N}} = \sum \frac{T_i}{\|T\| \sqrt{N}}$$

(6)

Equation (6) helps to find $\|T\| \cos^2 \theta_T$ from the data, even if $\|T\|$ or $\cos^2 \theta_T$ are not known.

Equating (3) and (5), we get

$$\sum \frac{X_i}{\|X\| \sqrt{N}} = \cos^2 \theta_X \sec^2 \theta_T$$

(7)

which gives relationship between $\|X\|$ and $\|T\|$.
Let $\theta_{E}$ represents the angle between the error score vector and the ideal vector. As per classical test theory

$$E = 0 \quad \text{i.e.,} \quad E = \|E\| \cos \theta_{E} = 0 \Rightarrow \cos \theta_{E} = 0 \quad \text{since} \quad \|E\| \neq 0$$

(8)

**Variance**

Test variance or observed score variance can also be obtained from $N$-dimensional person space by $S_{X}^{2} = \frac{\|X\|^{2}}{N}$ where $x_{i} = X_{i} - \bar{X}$ represents deviation scores.

Alternatively, $S_{X}^{2} = \frac{\|X_{i}\|^{2}}{N} - \bar{X}^{2} \Rightarrow S_{X}^{2} = \frac{\|X_{i}\|^{2}}{N} - \frac{\|X\|^{2}}{N} \cos^{2} \theta_{X}$

by equation (3)

$$\Rightarrow S_{X}^{2} = \frac{\|X_{i}\|^{2}}{N} \sin^{2} \theta_{X}$$

(9)

Thus, standard deviation (SD) of test score is the product of the length of the score vector and sine of the angle between the score vector and the ideal vector divided by the sample size.

Thus, the relationship between $x_{i}$ and $X$ is given by

$$\|x_{i}\|^{2} = \|X_{i}\|^{2} \sin^{2} \theta_{X}$$

(10)

Similarly, true score variance is obtained

$$S_{T}^{2} = \frac{\|T\|^{2}}{N} - \frac{\|T\|^{2}}{N} \cos^{2} \theta_{T} = \frac{\|T\|^{2}}{N} - \frac{\|X\|^{2}}{N} \cos^{2} \theta_{X}$$

(11)

Equation (11) helps to find value of true score variance of a test from the data if the length of true score vector is known, and

$$\|T\|^{2} = \|X\|^{2} \sin^{2} \theta_{T}$$

(12)

Now $S_{X}^{2} = S_{T}^{2} + S_{E}^{2}$ implies $\|X\|^{2} \sin^{2} \theta_{X} = \|T\|^{2} + \|X\|^{2} \cos^{2} \theta_{X} + \|E\|^{2}$

$$\Rightarrow \|X\|^{2} = \|T\|^{2} + \|E\|^{2}$$

(13)

**Correlations**

One assumption of classical test theory is $r_{TE} = 0 \Rightarrow \text{Cov}(T, E) = 0$

$$\Rightarrow \sum T_{i} E_{i} = 0 \Rightarrow \sum T_{i} (X_{i} - \bar{T}) = 0$$

One may also write $\sum T_{i} E_{i} = 0 \Rightarrow \sum (X_{i} - \bar{E}) E_{i} = 0$

$$\Rightarrow \sum X_{i} E_{i} = \|E\|^{2} \Rightarrow \text{Cov}(X, E) = S_{E}^{2}$$

(14)

Thus

$$S_{X}^{2} = \text{Cov}(X, T) + \text{Cov}(X, E)$$

(15)

So variance–covariance matrix of observed score $(X)$, true score $(T)$, and error score $(E)$ of a test is

$$\begin{pmatrix}
S_{X}^{2} & S_{XT} & S_{XE} \\
S_{XT} & S_{T}^{2} & S_{TE} \\
S_{XE} & S_{TE} & S_{E}^{2}
\end{pmatrix}$$

and trace of the matrix is $2S_{X}^{2}$.

It is well known that correlation between $X$ and $T$ is given by $r_{XT} = \cos \theta_{xt} = \frac{\sum x_{i} t_{i}}{\|X\| \|T\|}$, where $x_{i}$ and $t_{i}$ denote deviation scores and $\theta_{xt}$ is the angle between two deviation score vectors.

Now, $\sum x_{i} t_{i} = \sum (X_{i} - \bar{X})(T_{i} - \bar{T}) = \sum (X_{i} T_{i} - \bar{X} \bar{T}) - \sum (X_{i} \bar{T} + \bar{T} X_{i} - \bar{X} \bar{T})$

$$\Rightarrow \sum X_{i} T_{i} = \|T\|^{2} - N \bar{T}^{2}$$

(since, $r_{TE} = 0$)

$$\sum X_{i} T_{i} = \|T\|^{2} - N \bar{T}^{2}$$

(16)

But $\cos \theta_{xt} = \frac{\sum X_{i} T_{i}}{\|X\| \|T\|} = \frac{\|T\|^{2}}{\|X\| \|T\|}$ (since $r_{TE} = 0$)

$$\Rightarrow \sum X_{i} T_{i} = \|T\|^{2}$$

Therefore

$$\cos \theta_{xt} = \frac{S_{T}^{2}}{S_{X}^{2} \|X\|^2}$$

(17)

Equation (17) gives the relationship between $\cos \theta_{xt}$ (which is equal to $r_{XT}$) and $\cos \theta_{XT}$.

Note that
Value of $\cos \theta_{gh}^2$ can be obtained from the data using equation (3). Value of $\sin \theta_{gh}^2$ can also be obtained accordingly or by using equation (10).

Value of $\|\mathbf{X}_g\|^2$ which is equal to $\|\mathbf{X}_h\|^2$ can be obtained from data even if the value of $\|\mathbf{X}_g\|$ or $\sin \theta_{gh}$ is unknown.

Equations (15) to (17) each involve unknown quantity and thus does not help to compute directly $r_n$ or $S_r^2$ or $\|\mathbf{E}\|^2$.

Reliability and parallel tests

A test consisting of $n$-items when dichotomized in parallel halves say $g$th subtest and $h$th subtest results in two points $\mathbf{X}_g$ and $\mathbf{X}_h$ in the $N$-dimensional person space. As per classical definition, two tests “$g$” and “$h$” are parallel if $T_{i}^{(g)} = T_{i}^{(h)}$ and $S_{i}^{(g)} = S_{i}^{(h)}$, where the subscript “$g$” refers to subtest $g$ and subscript “$h$” refers to subtest $h$, and $S_{i}^{(p)}$ is the SD of error scores of the $p$th subtest, $p = g, h$. It can be proved that if $g$ and $h$ are parallel, $\mathbf{X}_g = \mathbf{X}_h$ and $S_{gh} = S_{gh}$.

Also, $\mathbf{X}_g = T_g + E_g$ and $\mathbf{X}_h = T_h + E_h$. Now $T_{i}^{(g)} = T_{i}^{(h)}$ implies $X_g - X_h = E_g - E_h$, so that

$$
\|\mathbf{X}_g\|^2 + \|\mathbf{X}_h\|^2 - 2\|\mathbf{X}_g\|\|\mathbf{X}_h\|\cos \theta_{gh} = 
E_g^2 + E_h^2 - 2E_g\|\mathbf{E}_h\|\cos \theta_{gh}^2
$$

(18)

where $\theta_{gh}$ is the angle between $\mathbf{X}_g$ and $\mathbf{X}_h$ while $\theta_{gh}^2$ is the angle between $E_g$ and $E_h$. But correlation between error scores of two parallel tests is zero. Thus

$$
\|\mathbf{X}_g\|^2 + \|\mathbf{X}_h\|^2 - 2\|\mathbf{X}_g\|\|\mathbf{X}_h\|\cos \theta_{gh} = 
E_g^2 + E_h^2 - 2N\Sigma^2
$$

(19)

since $S_{gh}^2 = S_{gh}^2$ by definition. The above equation suggests

$$
S_{gh}^2 = \frac{1}{N}\underbrace{\|\mathbf{X}_g\|^2 + \|\mathbf{X}_h\|^2 - 2\|\mathbf{X}_g\|\|\mathbf{X}_h\|\cos \theta_{gh}}_{2N\Sigma^2}
$$

(20)

Hence

$$
r_n = \frac{1}{N}\underbrace{\|\mathbf{X}_g\|^2 + \|\mathbf{X}_h\|^2 - 2\|\mathbf{X}_g\|\|\mathbf{X}_h\|\cos \theta_{gh}}_{N\Sigma^2}
$$

(21)

Equation (20) helps to find value of error variance of the test and hence true score variance as $(S_X^2 - S_r^2)$, and use them directly in equation (21) to find reliability of the test as per the classical definition from a single administration in terms of length of score vectors of two parallel tests and angle between such vectors. Thus, it is possible to find true score variance from the data and to calculate a reliability that conforms to the theoretical definition even if true scores of individuals taking the test are not known.

Note that while equation (21) finds reliability ($r_n$) of a test which is isomorphic to definition, the process of splitting the test in two parallel halves gives the popular method of finding split-half reliability of a test as correlation between two parallel forms of the test ($r_{gh}$). However, value of $r_{gh}$ may be different from value of $r_n$.

The approach of finding reliability as per definition can be extended to find reliability of a battery of tests aiming to measure a finite number of variables. After administration of the battery to $N$-individuals, values of $S_X^2$, $S_E^2$, $S_r^2$, and $r_n$ for each constituent test can be computed. Method of obtaining reliability of the battery depends on these parameters and also on definition of battery score.

If battery score is taken as sum of score of $K$-tests (summative score), it can be proved that

$$
r_{n(battery)} = \sum_{i=1}^{K} r_{i}(i)S_X^2 + \sum_{i=1}^{K} \sum_{j=1}^{K} 2Cov(X_i, X_j)
$$

$$
\sum_{i=1}^{K} \sum_{j=1}^{K} \Sigma_{i}^{2} + \sum_{i=1}^{K} \sum_{j=1}^{K} 2Cov(X_i, X_j)
$$

(22)

If $W_i, W_2,...,W_K$ are the weights to $K$-constituent tests of a battery where $W_i \geq 0 \ \forall i = 1, 2, 3,..., K$ and $\sum W_i = 1$ and the battery score be $Y_i = \sum_{i=1}^{K} W_i X_i$.

Clearly, $\text{var}(Y) = \sum_{i=1}^{K} W_i^2 \text{var}(X_i)$. Reliability of the battery can be computed as

$$
r_{n(battery)} = \sum_{i=1}^{K} r_{i(i)}W_i^2S_X^2 + \sum_{i=1}^{K} \sum_{j=1}^{K} 2W_iW_jCov(X_i, X_j)
$$

$$
\sum_{i=1}^{K} W_i^2S_X^2 + \sum_{i=1}^{K} \sum_{j=1}^{K} 2W_iW_jCov(X_i, X_j)
$$

(23)

Validity

Usual procedure of obtaining validity of a test is to find correlation of the test score ($X$) and criterion score ($Y$). It is known that $r_{XY} = \cos \theta_{XY}$, where $\theta_{XY}$ is the angle between the vectors of deviation scores. Linear regression of $Y$ on $X$ can also be represented by the vector resulting from the projection is $\hat{y} = p(y \ on \ x)$, the vector of predicted values on $Y$.

Item statistics for test consisting of binary items

Difficulty value of a test

If $X$ coincides with $I$, then the test is extremely easy since each subject has got maximum possible score. Thus,
difficulty value of a test should consider twofold criteria, namely, \( \theta \) and ratio of \( ||X|| \) and \( ||X|| \).

Accordingly, difficulty value of a test \( D \) may be defined as
\[
D = \frac{||X|| \cos \theta}{||I||} = \frac{X}{n} \tag{24}
\]

Note that equation (24) defines difficulty value of a test as ratio of product of length of the observed score vector and cosine of the angle between the observed score vector and the ideal vector as the numerator and length of the ideal vector as the denominator, keeping harmony with the usual notion of difficulty value of a test which actually measures degree of easiness of a test. It can be proved that \( 0 \leq D \leq 1 \).

**Item difficulty value**

In the \( N \)-dimensional person space, scores of an item may be characterized by an \( N \)-dimensional vector whose components are 0s and 1s. For an item, the vector \( I \) represents maximum possible score with components \( I_1, I_2, \ldots, I_N \), where \( I_i = 1 \) for all \( i = 1, 2, \ldots, N \). Thus, \( ||I|| = \sqrt{N} \). If \( K \)-persons (\( K \leq N \)) answer the \( i \)th item correctly, then \( ||X|| = \sqrt{K} \). Thus, difficulty value of the \( i \)th item is defined by
\[
D_i = \frac{||X||^2}{||I||^2} = \frac{K}{N} \tag{25}
\]

Clearly, \( 0 \leq D_i \leq 1 \). It may be observed that difficulty value of an item as per equation (25) in terms of ratio of square of length of the observed score vector and square of length of the ideal vector coincides with normal idea of proportion of persons passing an item.

**Discriminating value of a test**

If the vector \( X \) makes a zero degree angle with the vector \( I \), then the test fails to discriminate the subjects. So \( \theta \) or a suitably defined function of \( \theta \) will measure the discriminating value of a test. Since it is desirable for the discriminating value to lie in \([0, 1]\), \( \tan \theta \) will measure the discriminating value of a test. Thus
\[
D^i = \tan \theta = \frac{S^i}{X} \quad \text{(from equations (1) and (9))} \tag{26}
\]

where \( D^i \) denotes the discriminating value of a test.

Thus, discriminating value of a test equals to tan of the angle between observed score vector and ideal vector which is the ratio of \( SD \) and mean of the test score. In other words, discriminating value of a test is a linear function of its coefficient of variation (CV), a well-known measure of relative precision which is independent of change of scale but not of origin.

**Discriminating value of an item**

The method also helps to find discriminating value of an item as follows
\[
D^i = \tan \theta_{Xi} \tag{27}
\]

where \( D^i \) denotes the discriminating value of the \( i \)th item and \( \theta_{Xi} \) is the angle between the observed score vector for the \( i \)th item and ideal vector of an item with each component equal to unity.

Note that this method of finding discriminating value of an item does not consider total test score. The method considers entire data relating to the item and avoids consideration of top 27% and bottom 27% of data. However, item discriminating value tends to be lower for tests measuring different content areas and cognitive skills.

**Evaluation of \( \tan \theta \)**

For the \( i \)th item, the components of a vector \( X \) are \( K \)-numbers of 1s and rest 0s, if \( K \)-persons could answer the item correctly, where \( K \leq N \). Each of the \( N \)-component of the ideal vector is equal to unity. Let \( \theta_{Xi} \) denotes the angle between the vectors \( X \) and \( I \). Here
\[
\cos \theta_{Xi} = \sqrt{\frac{K}{N}} \Rightarrow \tan \theta_{Xi} = \sqrt{\frac{N-K}{K}} \tag{28}
\]

Note that

1. The discriminating value of an item is equal to the ratio of \( SD \) of the item score and mean of the item score, and can exceed unity. It is a function of CV.
2. If \( K=0 \), that is, if all the subjects fail in an item, then discriminating value is not defined for the item.
3. If \( K=N \), that is, if all the subjects pass an item, then discriminatory value is zero for that item.
4. It can be shown that square of discriminating value of an item is equal to reciprocal of difficulty value minus one. That is
\[
\left[ D^i \right]^2 = \frac{1}{D^i} - 1 \tag{29}
\]

Equation (29) gives the non-linear relationship between difficulty value and discriminating value of items.

5. Relationship between difficulty value and discriminating value of a test can be found as follows
Thus, product of difficulty value and discriminating value of a test is equal to $SD$ of the test divided by the number of items.

**Relationships between test reliability and discriminating value of a test**

Since $\bar{X} = \bar{T}$ and test reliability $r_{st} = S_T^2 / S_X^2$

$$r_{st} \left(D^1\right)^2 = \frac{S_T^2}{S_X^2} = \left(\frac{S_T}{\bar{X}}\right)^2 = \left(\frac{S_T}{\bar{T}}\right)^2$$

Thus, product of test reliability and square of test discriminating value is equal to square of CV of true scores. However, verification of the relationship may require finding the discriminating value of a test.

**Item-total correlations**

Point-biserial correlation coefficient ($r_{pb}$) is the proper statistic to estimate the degree of relationship between score of an item (dichotomous variable) and test score (interval/ratio scale) and is defined as $r_{pb} = \frac{(M_p - M_q) \sqrt{pq}}{S_Y}$, where $r_{pb}$ denotes point-biserial correlation coefficient, $M_p$ denotes test mean for persons answering the item correctly (i.e. those coded as 1s), $M_q$ denotes test mean for persons answering the item incorrectly (i.e. those coded as 0s), $S_Y$ denotes $SD$ of the test scores, $p$ denotes proportion of persons answering correctly (i.e. those coded as 1s), and $q$ denotes proportion of persons answering incorrectly (i.e. those coded as 0s).

Let $X_s$ be the vector showing scores of the $s$th item in the $N$-dimensional person space. If $K$-persons could answer the $s$th item correctly, then $\|X_s\|^2 = K$.

Now, $X_s^t X_s = \sum \sum X_s X_{st} = $ sum of test scores for persons answering the $s$th item incorrectly = sum of test scores = $\sum \sum X_s X_{st} = \sqrt{N} \|X_s\| \cos \theta_X - \sum \sum X_s X_{st}$ (using equation (1)).

Here

whole-test mean for persons

$$M_{ps} = \frac{\sum \sum X_s X_{st}}{K}$$

whole-test mean for persons

$$M_{qs} = \frac{\sum \sum X_s X_{st}}{N - K}$$

$SD$ of the whole test $S_{\theta} = \frac{\|X_s\| \sin \theta_X}{\sqrt{N}}$

$\rho_s = $ Proportion of persons answering the $s$th item correctly = $\frac{K}{N}$

$q_s = $ Proportion of persons answering the $s$th item incorrectly = $\frac{N - K}{N}$
Thus, point-biserial correlation between the \( i \)th item \( r_{pb(i)} \) and test score may be obtained by

\[
    r_{pb(i)} = \frac{\sum \sum X_i X_{ij} - \frac{1}{K} \sum \sum X_i \sum X_{ij}}{\sqrt{\frac{N}{N-K}} X \| \cos \theta_X - \sum \sum X_i X_{is} \| \sqrt{N}}
\]

\[
    p_{pb} = \frac{\sum \sum X_i X_{ij} - \frac{1}{K} \sum \sum X_i \sum X_{ij}}{\sqrt{\frac{N}{N-K}} X \| \cos \theta_X - \sum \sum X_i X_{is} \| \sqrt{N}}
\]

(32)

This may be simplified further as

\[
    r_{pb(i)} = \frac{X^T X_i - \sum X_i X_{ij}}{\sqrt{N S_X}}
\]

\[
    p_{pb} = \frac{M_p - M_q}{\sqrt{N S_X}}
\]

(33)

**Item reliability**

In addition to item difficulty and item discrimination, the concept of item reliability is also used in item statistics. Item reliability is defined as a product of SD of item scores and a correlational discrimination index, usually item-total correlation. Thus, reliability of the \( i \)th item could be taken as

\[
    S_i r_{pb(i)} = \frac{\sqrt{N (M_p - M_q) p q}}{S_X} = \frac{(N-K) (M_p - M_q)}{S_X}
\]

(34)

where \( S_X = \sqrt{N p q} \) denotes SD of the score of \( i \)th item and \( r_{pb(i)} \) denotes point-biserial correlation coefficient of the \( i \)th item, and the number of persons answering correctly the \( i \)th item is \( K \). However, \( \cos \theta_X \), that is, the angle between score vector of the \( i \)th item and total score vector of the test \( X \), indicates a better measure of reliability of the \( i \)th item.

Higher values of item reliability are desirable. If two items have equal item discriminating value, then the item with higher variance is preferred to be retained.

**Likert-type test**

**Background**

Suppose \( n \)-respondents have answered each of the \( m \)-items of a Likert-type questionnaire, where each item has \( k \)-numbers of response categories, marked as 1, 2, ..., \( k \). Let \( X_{ij} \) be a general element of the basic data matrix where \( n \)-individuals are in rows and \( m \)-items are in columns. \( X_{ij} \) represents score of the \( i \)th individual for the \( j \)th item, where \( 1 \leq X_{ij} \leq k \).

Note \( \sum_{j=1}^{m} X_{ij} = \text{Sum of scores of all individuals for the \( j \)th item} \); \( \sum_{i=1}^{m} X_{ij} = \sum_{j=1}^{m} \text{Sum of scores of all the items for \( i \)th individual} \), that is, total score of the \( i \)th individual(individual score); and \( \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij} = \sum_{j=1}^{m} \text{Sum of scores of all the individuals on all the items} \), that is, total test score. In addition, one can have another matrix \( (f_{ij}) \) of order \( m \times k \) showing frequency \( i \)-th item to \( j \)-th response category. A row total = the sample size (\( n \)). Similarly, a column total will indicate total number of times that response category was chosen by all the respondents. Denote the column total of \( j \)th response category by \( f_{0j} \) for \( j=1,2,\ldots,k \). Clearly, \( \sum_{j=1}^{k} f_{0j} = \text{grand total} = (\text{sample size}) \times (\text{number of items}) = mn \).

After administration of the questionnaire to \( n \)-respondents, we can calculate empirical probabilities of each response categories of the \( i \)th item as \( p_{ij} = (p_{1j}, p_{2j}, \ldots, p_{kj})^T \). Note that \( \sum_{j=1}^{k} p_{ij} = 1 \). The \( i \)th item will fail to discriminate or it will have zero discriminating value if all respondents choose the \( j \)th response category, that is, \( p_{ij} = 1 \) or if \( p_{ij} = 0 \). Discriminating value of the \( i \)th item may be found as a measure of dissimilarities between these two vectors \( P_i \) and \( P_{0i} \), where \( P_{0i} \) represents the vector of equal probabilities. Similarly, for the scale, vector showing empirical probabilities will be \( T = \left( \frac{f_{01}}{mn}, \frac{f_{02}}{mn}, \ldots, \frac{f_{0k}}{mn} \right)^T \).

Clearly, \( \sum_{j=1}^{k} f_{0j} = \text{grand total} = (\text{sample size}) \times (\text{number of items}) = mn \).

Reliability of an item can be measured as cosine similarity between \( P_i \) and \( T \). While reliability is a measure of association of two vectors, discrimination is a measure of dissimilarities between the vectors. Thus, a negative relationship is expected between reliability and discriminating value. Computation of the vectors \( P_i \)'s, \( P_{0i} \), and \( T \) forms the stage to obtain discriminating values and reliability of Likert-type scale and items from the data.

**Measure for reliability of Likert-type test**

Chakrabartty (2018) proposed two nonparametric methods of finding reliability of Likert-type scale and items considering only the frequency or proportion for each cell of the item-response category matrix without involving any assumptions of continuous nature or linearity or normality for the observed variables, or the underlying variable being measured.

Cosine similarity between \( i \)th and \( j \)th item is

\[
    \cos \theta_{ij} = \frac{P_i^T P_j}{\| P_i \| \| P_j \|}
\]

(35)
where $\theta_{ij}$ is the angle between the vectors $P_i$ and $P_j$.

Item reliability in terms of correlation between the $i$th item score and total score can be obtained by

$$\cos \theta_{ij} = \frac{P_i^T T}{\|P_i\| \|T\|} \quad (36)$$

where $\theta_{ij}$ is the angle between the vectors $P_i$ and $T$.

Test reliability with $m$-items was found by replacing the polychoric correlation between items $i$ and $j$ by $\cos \theta_{ij}$

$$r_{st} = \frac{m}{m-1} \left(1 - \frac{1}{m + \sum \cos \theta_{ij}}\right) \quad (37)$$

For Bhattacharyya’s similarity, each vector $P_i$ and $P_j$ was converted to unit vector $\pi_i$ and $\pi_j$, where $\pi_i = \sqrt{P_i / \|P_i\|}$

and $\pi_j = \sqrt{P_j / \|P_j\|}$ so that $\|\pi_i\|^2 = \|\pi_j\|^2 = 1$. Here, item reliability in terms of item-test correlation was computed from

$$\rho(\pi_i, \sqrt{T}) = \sum_{j=1}^{k} \frac{f_{ij}}{mn} p_{ij} \quad (38)$$

and test reliability was taken as

$$\cos(\Theta) = \cos \left[ \cot^{-1} \frac{\sum \cos \Theta_i}{\sum \sin \Theta_i} \right] \quad (39)$$

where $\Theta$ denotes mean or most preferred direction among the angles $\Theta_1, \Theta_2, \Theta_3, \ldots, \Theta_k$ between unit vectors.

Each method enables to find reliability irrespective of distribution of the observed or underlying variables and avoiding test of unidimensionality or assumption of normality for Cronbach’s alpha or bivariate normality for polychoric correlations. Bhattacharyya’s similarity helps to find range of reliability of the entire Likert-type scale.

**Discriminating value of Likert-type item and Likert-type scale**

The concept of CV as a measure of discriminating value of a binary item can be extended to find discriminating value of a Likert-type item with multi-response alternatives and define $D_i = SD_i / \text{Mean}_i$ for the $i$th item. Discriminating value of the test is defined accordingly, that is, $D_{\text{test}} = \frac{SD_{\text{test}}}{\text{Mean}_{\text{test}}}$. Note that mean of the $i$th item is $\frac{1}{n} \sum X_{ij} = f_{0i}/n$ and variance of the $i$th item is $n \|P_i\|^2 - \left(\frac{f_{0i}}{n}\right)^2$.

If $\theta_X$ be the angle between the test score vector and the maximum possible vector $\mathbf{1}$ (each component is equal to $k$ for $k$-point scale), it can be proved that

$$\frac{CV_{\text{test}}}{\text{Mean}_{\text{test}}} = \|X\| \frac{\sin \theta_X}{\|X\| \cos \theta_X} = \tan \theta_X \quad (40)$$

CV is a well-known measure of relative precision which is independent of change of scale but not of origin. Note that $SD / \text{Mean}$ is:

- non-negative if response categories are marked with positive numbers;
- defined even if $p_{ij} = 0$ for a particular response category of $i$th item;
- $CV = 0 \iff$ either $p_{ij} = \frac{1}{k} \forall j = 1, 2, \ldots, k$ or $p_{ij} = 1$ for a particular $j$th category;
- $D_i = D_i^* \iff P_i = P_j^*$;
- possible to express $D_{\text{test}}$ as a function of $D_i^*$ since $\text{Mean}_{\text{test}}$ and $SD_{\text{test}}$ are combined mean and combined $SD$, respectively;
- shown that estimate of population CV is $\sigma / \mu$, where $\mu$ and $\sigma$ are population estimates of mean and $SD$, respectively;
- shown that statistical inferences for CV for normally distributed data are often based on McKay’s $\chi^2$ approximation for the CV (Forkman, 2013).

**Findings and conclusion**

Through $N$-dimensional person space, measures of test parameters and item statistics including difficulty value, discriminating value of test, Phi coefficient to depict correlations between a pair of items, and point-biserial correlation to depict item-total correlations with binary items were presented using angular similarity between two vectors and considering the entire data. Non-linear relationships between difficulty value and discriminating value of items and test were derived, including relationship between test reliability and test discriminating value.

Dichotomization of the test in parallel halves helps to find value of error variance, true score variance, and reliability of a test as per theoretical definition from a single administration in terms of length of score vectors of two parallel tests and angle between such vectors. The method was extended to find reliability of a battery of tests.

Reliability and discriminating value of a Likert-type scale and Likert-type item were found in terms of angular similarity using only the permissible operations for a Likert-type scale, that is, frequencies or probabilities of item-response categories without involving assumptions of continuous nature or linearity or normality for the observed variables, or the underlying variable being measured. Thus, such reliabilities and discriminating value are in fact nonparametric. The proposed methods also
avoid test of unidimensionality or assumption of normality or
bivariate normality associated with the polychoric correlations.
Thus, the proposed methods are considered as improvement
over the existing ones.

Reliability as a measure of association of two vectors and
discrimination as a measure of distance between the vectors
are likely to show a negative relationship. Further studies may
be undertaken to facilitate comparison of the proposed
methods.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect
to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, author-
ship, and/or publication of this article.

ORCID iD
Satyendra Nath Chakrabartty https://orcid.org/0000-0002-7687-5044

References
Chakrabartty SN (2018) Cosine similarity approaches to reliability
of Likert scale and items. *Romanian Journal of Psychological
Studies* 6(1): 3–16.

Chong HY, Sandra A, David W, et al. (2002) Teaching factor analy-
sis in terms of variable space and subject space using multi-
media visualisation. *Journal of Statistics Education*. Epub ahead
of print 1 December. DOI: 10.1080/10691898.2002.11910546.

Forkman J (2013) Estimator and tests for common coefficients of
variation in normal distributions. *Communications in Statistics
38*(2): 233–251.

Glass GV and Collins JR (1970) Geometric proof of the restriction
on the possible values of rXY when rXZ and rYZ are fixed.
*Educational and Psychological Measurement 30*: 37–39.

Huberty C (1994) Why multivariate analyses? *Educational and
Psychological Measurement 54*: 620–627.

Reyment R and Joreskog KG (1993) *Applied Factor Analysis in the
Natural Sciences*. Cambridge: Cambridge University Press.

Author biography
Satyendra Nath Chakrabartty is an M.Stat. from Indian Statistical
Institute. He was a research scholar at Psychometric Research and
Service Unit of Indian Statistical Institute. He has taught postgradu-
ate courses at Indian Statistical Institute, University of Calcutta,
Galgotias Business School, and so on. He has 50 publications to his
credit, including presentations in national and international seminars.
After serving Kolkata Port Trust for 25 years in various managerial
positions, he joined Mumbai Port Trust as director (Planning and
Research) and subsequently took over as director, Indian Institute of
Port Management, and retired from the position of director, Kolkata
Campus of the Indian Maritime University. Currently, he is associ-
ated with Indian Ports Association, New Delhi, as a consultant.