International Journal of Current Research in Science and Technology
Volume 9, Issue 02 February (2023)
ISSN: 2394-5745
https://doi.org/10.57181/ijcrst/vol9i01/57
Available Online: www.ijcrst.in

International Journal of Current Research in Science and Technology

Limitations of Neuropsychological tests and Remedial Measures

Research Article

Satyendra Nath Chakrabartty

Indian Statistical Institute, Indian Maritime University, Indian Ports Association

Coressponding Author: *Satyendra Nath Chakrabartty,* ***Email:*** *chakrabarttysatyendra3139@gmail.com*

Abstract:

Background: Neuropsychological tests differ in terms of length (number of items), width (number of response-categories/levels), sub-classes or dimensions covered, psychometric documentation for unambiguous interpretations, responsiveness, discriminating value, etc. and scores are not comparable.
Aim: To address methodological issues of neuropsychological testing and suggests remedial measures by transforming discrete item scores to continuous scores in ratio scale following normal distribution for meaningful evaluation of measurement properties and better utilization of such tests.
Methods: Using data driven weights to response-categories of different items, ordinal item scores are converted to equidistant score (E-scores) in ratio scale with fixed zero point. Proposed scores (P-scores) obtained from E-scores via standardization and linear transformation follows normal distribution, parameters of which can be obtained from data. Sub-class scores and test scores are obtained as sum of item-wise P-scores.
Results: Normally distributed P-scores facilitate aggregation with cardinal measures like number of errors, time taken, etc. and offer platform for parametric analysis including statistical testing. In addition, the proposed method helps to find reliability as per theoretical definition, factorial validity avoiding criterion variable, discriminating value, assessment of progress/deterioration of one or a group of patients, efficiency of classification, equivalent scores of two neuropsychological tests, etc.
Conclusion: Proposed scores following normal distribution and satisfying desired properties of measurement is recommended. Practicing psychiatrists and researchers can derive benefits of the proposed score for meaningful comparisons, classifications, equating premorbid scores, and assessment of progress or deterioration.

Keywords: Equivalent scores; Factorial validity; Neuropsychological Test Battery; Normal distribution; Theoretical reliability; Factorial validity.

INTRODUCTION

Measurement issues and associated statistics including psychometrics are foundational elements in neuropsychological tests. Neuropsychological tests are used for diagnostic purpose and treatment insights for mental health disorders. Neuropsychological tests involve *K*-point Likert/Numeric Rating scales (NRS) marked as 1, 2, 3, …., *K* pertaining to sub-classes or dimensions like Motor function (walking, coordination, etc.), Perception (how well the patients take in what they see or read), Problem-solving and decision-making, Verbal ability, Memory, Intelligence, Executive Functions, Language, Visuo-spatial, Multiple Functions, etc. For example, in the Halstead-Reitan Battery (HRB), participants decide whether a geometric figure best corresponds to the number 1, 2, 3, or 4.

Aggregating ordinal scores with cardinal measures like number of errors (Seashore Rhythm Test of HRB), Time taken to complete (Tactual Performance Test of HRB) etc. are problematic. Depending on patient symptoms, neuropsychologists decide the tests for assessment of the patient's cognitive abilities/disorders, better understanding of current health picture, medical needs and also to know whether the problems are due to Traumatic brain injury, Infections of brain and spinal cord,

Normal brain changes with age, or diseases like Alzheimer's disease, Dementia, Parkinson's disease and other disorders.

Large numbers of Neuropsychological tests are there to evaluate cognitive impairments. Choices of the tests are usually made considering relevant dimensions and quality measures like reliability and validity. However, no test uses theoretical definition of reliability. Reliability by Test-retest, Cronbach alpha and validity as correlation with criterion variable have inherent problems. Other quality measures like Discriminating value, responsiveness (ability to assess improvement or deterioration), etc. could also be critical in selection of tests. Reliability, validity, for sub-scale and test are different for *k*-point scales, *k*= 2, 3, 4, 5, ... etc.

The paper highlights methodological issues of neuropsychological testing and suggests remedial measures by transforming discrete item scores to continuous, monotonic scores following normal distribution for meaningful evaluation of properties and better utilization of such tests.

Problem Areas

Neuropsychological tests with different item formats (length and number of response-categories/levels), sub-classes/dimensions covered, scoring methods, etc. are not comparable. Reporting of results by mean ± standard deviation (SD) assuming admissibility of addition of ordinal data, equidistant response-categories and comparing means by *t*-test, analysis of variance (ANOVA), independent factors by Principal component analysis (PCA), Factor analysis (FA), assuming normal distribution of scores, etc. are unjustified.

Neuropsychological tests have been criticized from insides [1] and outside the discipline [2]. Neuropsychological Test Battery (HRNB) lacks basic psychometric documentations to facilitate interpretations of result [3]. Methodological errors in construction of Luria–Nebraska Neuropsychological Battery (LNNB) are significant [4]. Similar problems exist for normative data of Benton Test of Facial Recognition, Mirsky's Continuous Performance Test, National Adult Reading Test, Purdue Pegboard Test, Rey Complex Figures, Stroop Color and Word Test, etc. and are far below contemporary standards [5].

Nature of Data

Data generated by NRS/Likert items consist of frequency in each response-category giving rank order information. Numbers assigned to response-categories are a way to provide ranking responses. Levels are ordered but not equidistant. If $d_{j(j+1)}$ denotes distance between *j*-th and *(j+1)*-th levels of an item, equidistant property demands constant $d_{j(J+1)}$ $\forall j$ =1, 2, 3, 4 for a 5-point item.

Rating data are often skewed, have floor or ceiling effects, and require normality checks for inferences [6]. Discrete ordinal data are not normally distributed; violet assumptions of many statistical procedures [7] and result in problems for undertaking parametric statistical analysis [8]. A scale must have the following features: metric, presence of zero point, and clearly defined operational procedure as the basis for measurement [9].

Scoring

Arithmetic averages requiring equidistant scores are not meaningful for ordinal item scores [10]. Distances between response-categories are unknown and not uniform [11]. Test score as sum of independent dimensions amounts to adding apples with oranges. Equal importance to the items for summative score is not justified since items differ in contributions to total score, item-total correlations, and factor loadings [12]. Non-admissibility of addition implies mean, SD, correlation, regression, ANOVA, FA, PCA, Cronbach alpha using sum of item variances and test variance, etc. are not meaningful and may produce strange results [13]. Using parametric analysis with ordinal data assuming normal distribution is one of the seven deadly sins of statistical analysis [14].

Interval scales have constant, equal distances between values but the zero point is not fixed. Thus, difference between two measurements has meaning, but their ratio does not [15].

Responses to different response-categories of items can generate tied scores. Thus, summative scores fail to distinguish the respondents with tied scores and reduce discriminating value of the scale.

Distribution of Score

Neuropsychological tests do not consider distribution of scores. Item scores depend on endorsed response-categories and do not follow similar distributions. Unknown and different distributions of item scores make it difficult to interpret $X \pm Y$ and to find joint distribution of $X \pm Y$. Addition of two random variables $X + Y = Z$ is most meaningful if $P(Z = z) = P$ (X= x, Y= z - x) for discrete case and for continuous case, $P(Z \leq z) = P\ (X + Y \leq z) = \int_{-\infty}^{\infty}(\int_{-\infty}^{z} f_{X,Y}(x, t-x)dt)\ dx$. Thus, it is necessary to know probability density function (pdf) of X and Y and their convolution. Moreover, sum of two log-normally distributed variables cannot

be obtained as such and require complex Lie-Trotter operator splitting method [16]. Problems of parametric statistical analysis with ordinal, skewed NRS data with ceiling and floor effects were addressed [8].

Non-Satisfaction of Assumptions

Neuropsychological test scores may not satisfy assumptions of statistical techniques like PCA, FA, *t*-test, paired *t*-test, *F*-test, etc., which assume normal distribution of the variables under study. Results may go wrong if assumptions of the techniques used are violated. For example, high correlation between two variables X and Y is taken as linear relationship between X and Y and ordinary least-squares (OLS) regression of the form $Y = \alpha + \beta X + \epsilon$ is fitted. However, r_{XY} may be high even if Y is non-linearly related with X. If X takes integer values from 1 to 30, $r_{X,X^2} = 0.97$; $r_{X,X^3} = 0.92$; $r_{X,log_{10}^X} = 0.92$ despite each of X^2, X^3, log_{10}^X is non-linear function of X. Clearly, *linearity implies high correlation but not the converse*.

Linearity between X and Y can be tested by checking normality of error score $E = (Y - \hat{Y})$ or testing $H_0: S_E^2 = 0$ where $S_E^2 = \frac{1}{n}\sum(Y_i - \hat{Y}_i)^2$ denotes variance of error scores for sample size n. Error score of $Y = \alpha + \beta X + \epsilon$ for $Y = X^2$ or X^3 or log_{10}^X, did not follow normal and violated assumption of OLS. This is an example to show how violation of assumptions of statistical analysis may mislead the results.

Better is to transform ordinal scores of i-th item to continuous equidistant scores (E_i-scores) in ratio scale and transform E_i-scores to proposed scores (P_i-scores) following normal distribution such that $1 \leq P_i \leq 100$

Cut-off point and classification

Consider another example where $Y = \frac{1}{\sqrt{2\pi}}\, e^{\frac{-1}{2}X^2}$ for $0 \leq X \leq 3.9$. Here, $r_{XY} = 0.93$. But, if $-3.9 \leq X \leq 3.9$, $r_{XY} = 0.00036$. Thus, truncated values of either X or Y or both can affect the correlation significantly. The point is relevant to decide cut-off point (X_0) where individuals with scores $\leq X_0$ are without the disease and those with scores exceeding X_0 have the disease, assuming higher score implies higher dysfunctions. Samples with high proportion of persons without disease or patients with the disease will tend to truncate distribution of test score. Thus, X_0 obtained from study-specific populations may affect clinical heterogeneity and not allow comparisons between studies.

Diagnosis of neurological disorder is complex since many symptoms can happen in different combinations of different disorders. Moreover, many disorders don't have definitive causes, markers. Biomarkers have limitations too. A case study by [17] indicated missing of core clinical features of Dementia with Lewy Body (DLB) and negative indicative biomarkers, but neuropsychological tests and Positron emission tomography (PET) imaging provided crucial evidence for DLB even in early stages. After follow-up, core symptoms and biomarkers appeared in later stages of the disease. Testing of cerebrospinal fluid (CSF) by standard two-tier testing algorithm (STTA), enzyme immunoassays (EIAs), and/or immunoblots for diagnosis of central nervous system Lyme disease; culture of CSF for Lyme Borreliais are not recommended [18].

D**iagnostic and Statistical Manual of Mental Disorders** (DSM)-IV is a categorical classification system which is prototypes. In addition, classifications in several classes are made like mild, moderate, severe forms of a disorder. However, no diagnosis is confirmed simply as a function of the data from a neuropsychological assessment. In the case of dementia, for instance, multiple additional criteria are required to meet for which relevant information are obtained from other sources.

Reliability

Common reliability measures of neuropsychological tests are:

Test-retest reliability reflecting stability of scores by correlation between two administrations of the test at two different time points on the same sample with same testing conditions with no agreed choice of retest-time frames [19]. It does not consider rank-order stability of individuals and may fail to show the extent of agreement between two administrations. Test-retest reliability may be high despite rejection of hypothesis $H_0: \mu_{X\,Test} = \mu_{X\,Retest}$ by paired *t*-test [20]. Difference between correlation and agreement was demonstrated empirically [21].

Cronbach alpha assumes one-dimensional test i.e. all items measure the same construct. Violation of the assumption may bias the coefficient α [22] and distort the variance-covariance matrix, if distribution of observed responses is not symmetric [23]. However, there are instances of reporting alpha despite multi-factors emerged from PCA or FA. Against suggestion of two-factor solution (memory factor and visuo-spatial factor) for Repeatable Battery for Assessment of Neuropsychological Status (RBANS) with 12 sub-tests, five index scores and a total scale score, Cronbach alpha = 0.92 of RBANS was found [24]. Eight independent factors of expanded Halstead-Reitan Battery (eHRB) out of

over forty test measures was found by [25], despite caution about inferring complex mental abilities by FA or structural equation modeling [26]. Test reliability ≠ Average of sub-tests reliabilities. Internal reliability coefficients of the Wechsler adult intelligence scale–Fourth Edition (WAIS-IV) was 0.98 against reliability of 0.96, 0.94, 0.95 and 0.90 respectively for Verbal comprehension, Working memory, Perceptual reasoning and Processing speed [27].

Avoiding unidimensionality assumption, [28] proposed finding theoretical reliability by dichotomizing a test in two parallel subtests (g-th and h-th) and finding Error variance S_E^2 as

$$S_E^2 = \frac{1}{N}[\|X_g\|^2 + \|X_h\|^2 - 2\|X_g\|\|X_h\|Cos\theta_{gh} \quad (1)$$

where N is the sample size; $\|X_g\| = \sqrt{\sum_{i=1}^{N} X_{ig}^2}$ denotes length of the g-th vector; $\|X_h\|$ is defined similarly and θ_{gh} is the angle between the g-th and h-th vectors. Thus,

$$r_{tt} = \frac{S_T^2}{S_X^2} = 1 - \frac{S_E^2}{S_X^2} = 1 - \frac{\frac{1}{N}[\|X_g\|^2 + \|X_h\|^2 - 2\|X_g\|\|X_h\|Cos\theta_{gh}]}{NS_X^2} \quad (2)$$

Reliability of a battery consisting of K-subtests could be found in terms of sub-test reliabilities (without weights) by $r_{tt\,(Battery)} = \frac{\sum_{i=1}^{K} r_{tti} S_{X_i}^2 + \sum_{i=1}^{K}\sum_{i\neq j}\sum_{j-1}^{K} 2\,Cov(X_i,X_j)}{\sum_{i=1}^{K} S_{X_i}^2 + \sum_{i=1}^{K}\sum_{i\neq j}\sum_{j-1}^{K} 2\,Cov(X_i,X_j)}$ (3)

Validity

For two different instruments X and C, criterion validity (r_{XC}) also reflects validity of C. High r_{XC} ⟹instrument X is not required. In addition, the approach assumes similarity of latent variables being measured by X and C and requires administration of both X and C to the same sample. For high positive skew of the test score(X) or criterion score(C), r_{XC} gets reduced if the data contains predominantly high performers [29]. To avoid such problems, structural validity of normally distributed transformed scores by PCA was preferred [30].

METHOD AND RESULTS

Proposed Method

(1) Ensure levels of a k-point item to 1, 2, 3, …., k avoiding zero.

(2) Ensure each item ispositively related to the test score i.e. higher item score indicates higher level of disorder.

Transformations of item scores

Convert ordinal item score to equidistant score (E) using different weights to response- categories of different items so that $W_1, 2W_2, 3W_3, \ldots\ldots, KW_k$ forms an arithmetic progression and satisfy equidistant property by method suggested by [30], which is briefly described below for n-number of respondents and $k = 5$:

I: Find maximum frequency (f_{max}) and minimum frequency (f_{min}) of the levels of each item. Consider $\frac{f_{ij}}{n}$ as initial weights(ω_{ij}).

Arrange $\omega_{ij}'s$ so that $\omega_{i1} < \omega_{i2} < \omega_{i3} < \omega_{i4} < \omega_{i5}$ where $\omega_{i1} = \frac{f_{min}}{n}$ and $\omega_{i5} = \frac{f_{max}}{n}$.

Choose intermediate weight $W_{i1} = \omega_{i1}$ Find the common difference α so that

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

Other intermediate weights are: $W_{i2} = \frac{\omega_{i1} + \alpha}{2}$, $W_{i3} = \frac{\omega_{i1} + 2\alpha}{3}$; $W_{i4} = \frac{\omega_{i1} + 3\alpha}{4}$; and

$W_{i5} = \frac{\omega_{i1} + 4\alpha}{5}$. Get final weights $W_{ij(Final)} = \frac{W_{ij}}{\sum_{j=1}^{5} W_j}$ enabling $\sum W_{ij(Final)} = 1$ and

$j.W_{j(Final)} - (j-1).W_{(j-1)(Final)}$ = constant, value of which is different for different items.

II: Standardizing E-scores to Z-scores by $Z = \frac{E - \bar{E}}{SD(E)}$ $\sim N(0,1)$

III**:** Convert Z-scores to proposed scores (P-scores) by $P = (99)\left[\frac{Z_{ij} - Min_{Z_{ij}}}{Max_{Z_{ij}} - Min_{Z_{ij}}}\right] + 1$ so that $1 \leq P \leq 100$ and P follows normal. Sub-scale score and test score of an individual is taken as sum of normally distributed item-wise P-scores.

Properties

i) E-scores are in ratio scale where fixed zero point occurs when $f_{ij} = 0$ for j-th level of i-th item.

ii) The method can be used for items with different values of k including binary items.

iii) For the i-th item, $P_i \sim N(\mu_i, \sigma_i^2)$ where μ_i and σ_i^2 can be estimated from the data.

iv) P-scores for sub-class and test are continuous, monotonic, normally distributed with better admissibility of arithmetic aggregation and facilitating parametric analysis including testing of hypothesis like $H_0: \mu_1 = \mu_2$ or $H_0: \sigma_1^2 = \sigma_2^2$ etc. either for longitudinal data or snap-shot data.

v) Percentage progress/deterioration of the i-th person in two successive time-periods can be assessed by $\frac{P_{i(t)}-P_{i(t-1)}}{P_{i(t-1)}} \times 100$, reflecting responsiveness of the scale and effectiveness of treatment plan. $P_{i(t)} - P_{i(t-1)} > 0$ implies progress in t-th period over (t-1)-th period. Similarly, progress for a group of persons is reflected if $\overline{P_{i(t)}} > \overline{P_{i(t-1)}}$. Deterioration ($P_{i(t)} - P_{i(t-1)} < 0$) may be probed to find extent of deterioration in sub-class scores for possible corrective actions.

vi) Normality of P_i helps to test H_0: $\mu_{P_t} = \mu_{P_{(t-1)}}$ or H_0: $Progress_{(t+1)over\ t} = 0$ avoiding need to find minimal important difference (MID) of a scale or testing effectiveness of treatments/cares by H_0: $\mu_{P_{pre-group}} = \mu_{P_{post-group}}$ using paired t-test since pre-treatment group and post-treatment group are not independent.

vii) Progress/deterioration of a patient or a group of patients across time can be plotted to compare progress pattern i.e. response to treatments from the start.

Normally distributed P-scores also help to find psychometric properties of the scale in better fashion.

Factorial validity

Normality satisfies the assumptions of PCA enabling computation of factorial validity as ratio of the first eigenvalue to the sum of all eigenvalues i.e. Factorial validity = $\frac{\lambda_1}{\sum \lambda_i}$, where λ_1 is the highest eigenvalue associated with the first principal component reflecting the main factor for which the test was developed. Such factorial validity avoids the problems of construct validity involving administration of two tests and selection of criterion scale [31].

Reliability

Population estimates of item variance and scale are possible for normally distributed P-scores. Such estimates can be used to find Cronbach alpha at population level for a domain/sub-class with n-items as

$$\hat{\alpha} = \left(\frac{n}{n-1}\right)\left(1 - \frac{Sum\ of\ estimates\ of\ variance\ of\ items\ in\ the\ sub-class)}{Estimate\ of\ variance\ of\ the\ sub-class}\right) \quad (4)$$

Discriminating value

Discriminating value reflects ability of the test to distinguish between individuals with more or less severe disease. Discriminating value of i-th item ($Disc_i$) and test ($Disc_{Test}$) can be computed by Coefficient of variation (CV) where $Disc_i = \frac{SD_i}{mean_i}$ and $Disc_{Test} = \frac{SD_{Test}}{Mean_{Test}}$. For a test with m-items, variance of the i-th item $S_{X_i}^2 = \bar{X_i}^2 . Disc_i^2\ \forall\ i = 1, 2 \ldots m$

$\Rightarrow \sum_{i=1}^{m} S_{X_i}^2 = \sum_{i=1}^{m} \bar{X_i}^2 . Disc_i^2$ and Test variance $S_X^2 = \bar{X}^2 . {Disc_T}^2$.

Thus, $\alpha = \left(\frac{m}{m-1}\right)\left(1 - \frac{\sum_{i=1}^{m} \bar{X_i}^2 . Disc_i^2}{\bar{X}^2 . {Disc_T}^2}\right)$

and $(Disc_{Test})^2 = \frac{CV\ True\ scores^2}{r_{tt}}$ where $r_{tt} = \frac{S_T^2}{S_X^2}$ (6)

Clearly, test reliability and $Disc_{Test}$ are related by a negative non-linear relationship.

Classification

Classifications of individuals to a number of mutually exclusive classes involve deciding boundary points, so that members within a class are similar and members between classes are dissimilar. Quartile clustering of P-scores following normal distribution assigns equal probability to the quartiles Q_1, Q_2, Q_3, Q_4 i.e.

$$\int_1^{Q_1} f(x)dx = \int_{Q_1}^{Q_2} f(x)dx = \int_{Q_2}^{Q_3} f(x)dx = \int_{Q_3}^{Q_4} f(x)dx \quad (7)$$

If needed, docile clustering may be used to have 10 classes. However, each classification should be evaluated in terms of clinical meaningfulness.

Efficiency of classification may be assessed by Davies-Bouldin Index (DBI) [32]. For K-number of classes, DBI is computed by

$$DBI_K = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1\ (i\neq j)}^{K} Max\left[\frac{DiamC_i - DiamC_j}{\|C_i - C_j\|}\right]$$

where diameter of i-th class $DiamC_i = \sqrt{\frac{\sum_{x_i \in C_i} \|x_i - C_i\|^2}{n_i}}$

C_i: Centroid or mean of the i-th class; n_i: Number of members in the i-th class.

Upper limit of DBI is 1 and lower value implies better efficiency.

Equivalent scores

Different tests have different cut-off points. Hence, for two tests A and B, one needs to ensure that cut-off points X_{0A} and X_{0B} are equivalent ($X_{0A} \Leftrightarrow X_{0B}$). If scores of tests A and B are transformed to follow normal distributions, [33] suggested that transformed score T_{0A} corresponding to X_{0A} and T_{0B} corresponding to X_{0B} are equivalent if

$$\int_{-\infty}^{T_{0A}} f(x)dx = \int_{-\infty}^{T_{0B}} g(y)dy \quad (8)$$

where $f(X)$ and $g(Y)$ denote pdf of transformed scores of test *A* and test *B* respectively. Equation (8) can be solved using Standard Normal probability table and can also be used to integrate different neuropsychological tests i.e. to find score combinations $\{X_{01}, X_{02}\}$ for Test-1 and Test-2 respectively, such that for a given score of X_0 in Test-1, is equivalent to Y_0 in Test-2 if $\int_{-\infty}^{X_0} f(x)dx = \int_{-\infty}^{Y_0} g(y)dy$ and vice versa.

Limitations

The study considered availability of complete responses from each respondent to neuropsychological tests. In practice, responses of few respondents may be incomplete. Elimination of entire incomplete data reduces the sample size and statistical power.

CONCLUSIONS

Proposed equidistant scores (*E*-scores) using data driven weights to response-categories of different items are in ratio scale with fixed zero point. *P*-scores obtained from *E*-scores via standardization and linear transformations are continuous folloing normal distribution, parameters of which can be obtained from data.

Sub-class scores and test scores as sum of item-wise *P*-scores facilitate aggregation with cardinal measures like number of errors, time taken, etc. and offer platform for parametric analysis including statistical testing.

The proposed method also helps to find reliability as per theoretical definition, factorial validity avoiding criterion variable, discriminating value, assessment of progress/deterioration of one or a group of patients, efficiency of classification, equivalent scores of two neuropsychological tests, etc.

Practicing psychiatrists and researchers can derive benefits of the proposed score in ratio scale for better comparisons and prognosis. Simulation studies may be undertaken to evaluate merits of the proposed approach with multi datasets.

DECLARATIONS

Acknowledgement

None

Sources of Funding

No funds, grants, or other support was received

*Financial or non-financial interests**: Nil***

Availability of data and material

Nil

Approval of Ethics Committee

Not required for this methodological paper

Informed consent: No data collected from patients

REFERENCES

1. Prigatano, G. P. (2003): Challenging dogma in neuropsychology and related disciplines. *Archives of Clinical Neuropsychology*, 18, 811–825.
2. Reschly, D. & Gresham, F. M. (1989): Current neuropsychological diagnosis of learning problems: A leap of faith. In C. R. Reynolds, and E. Fletcher-Janzen (Eds.), *Handbook of clinical child neuropsychology*. New York: Plenum Press
3. Dean, R. S. (1985): Review of the Halstead–Reitan neuropsychological test battery. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook*. Lincoln: University of Nebraska Press.
4. Adams, R. L. (1985): Review of the Luria–Nebraska neuropsychological battery. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook*. Lincoln: University of Nebraska Press.
5. Reynolds, Cecil and Mason, Benjamin. (2009): Measurement and Statistical Problems in Neuropsychological Assessment of Children. 10.1007/978-0-387-78867-8_9.
6. Clason DL and Dormody TJ (1994): Analyzing data measured by individual Likert-type items. *J Agric Educ* 35(4); 31–35
7. Harwell, M. R. and Gatti, G. G. (2001): Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, 71, 105-131.
8. Šimkovic M. and Träuble B. (2019): Robustness of statistical methods when measure is affected by ceiling and/or floor effect. *PloS one*, 14(8): e0220889. https://doi.org/10.1371/journal.pone.0220889
9. Yusoff, Rohana and Janor, R. M. (2014): Generation of an Interval Metric Scale to

Measure Attitude, *SAGE Open*, 1-16, 10.1177/2158244013516768.

10. Lewin RJP, Thompson DR, Martin CR, Stuckey N, Devlen J, Michaelson S and Maguire P. (2002): Validation of the Cardiovascular Limitations and Symptoms Profile (CLASP) in chronic stable angina. *J Cardiopulm Rehabi* 22:184-191.

11. Munshi, Jamal (2014): A Method for Constructing Likert Scales. *SSRN Electronic Journal*. 10.2139/ssrn.2419366.

12. Parkin D, Rice N, Devlin N. (2010): Statistical analysis of EQ-5D profiles: does the use value sets bias inferences? *Med Decis Making*; 30(5);556–565.

13. Marcus-Roberts, H. M., & Roberts, F. S. (1987): Meaningless statistics. *Journal of Educational Statistics*. 12, 383-394.

14. Kuzon, W. M., Urbanchek, M. G., & McCabe, S. (1996): The seven deadly sins of statistical analysis. *Annals of Plastic Surgery*, 37, 265–272.

15. ampbell MJ, Machin D, Walters SJ. (2010): *Medical Statistics: A Textbook for the Health Sciences*, 4th edition. John Wiley and Sons, Chichester, England

16. Lo, C. F. (2012): The Sum and Difference of Two Lognormal Random Variables. *Journal of Applied Mathematics,* Article ID 838397,10.1155/2012/838397

17. Bouter Caroline, Hansen Niels, Timäus Charles, Wiltfang Jens, Lange Claudia (2020): Case Report: The Role of Neuropsychological Assessment and Imaging Biomarkers in the Early Diagnosis of Lewy Body Dementia in a Patient with Major Depression and Prolonged Alcohol and Benzodiazepine Dependence. *Frontiers in Psychiatry*, Vol. 11 https://doi.org/10.3389/fpsyt.2020.00684

18. Theel ES, Aguero-Rosenfeld ME, PrittB, Adem PV, Wormser GP. (2019): Limitations and confusing aspects of diagnostic testing for neurologic Lyme disease in the United States. *J Clin Microbiol* 57: e01406-18.10.1128/JCM.01406-18.

19. Chmielewski, M., and Watson, D. (2009): What is being assessed and why it matters: The impact of transient error on trait research. *Journal of Personality and Social Psychology,* 97(1), 10.1037/a0015618 19586248

20. Altman, D. G. & Bland, J. M. (1995): Statistics Notes: Absence of Evidence is not Evidence of Absence. *British Medical Journal,* 311(7003), 485-485. 10.1136/bmj.311.7003.485.

21. Streiner, D. L., Norman, G. R., & Cairney, J. (2014): *Health measurement scales: a practical guide to their development and use*. Oxford University Press, UK.

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

23. Flora, David B. and Curran, Patrick (2004): An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data**,** *Psychol. Methods*; 9(4): 466–491.10.1037/1082-989X.9.4.466

24. De la Torre GG, Suárez-Llorens A, Caballero FJ, Ramallo MA, Randolph C, Lleó A, Sala I, Sánchez B. (2014): Norms and reliability for the Spanish version of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) Form A. *J Clin Exp Neuropsychol*. 36(10):1023-30. 10.1080/13803395.2014.965664

25. Patt, V. M., Brown, G. G., Thomas, M. L., Roesch, S. C., Taylor, M. J., & Heaton, R. K. (2018): Factor Analysis of an Expanded Halstead-Reitan Battery and the Structure of Neurocognition. *Archives of clinical neuropsychology: 33*(1), 79–101. https://doi.org/10.1093/arclin/acx043

26. McFarland, D. (2017): Evaluation of multidimensional models of WAIS-IV subtest performance. *The Clinical Neuropsychologist,* 31, 1127–1140. 10.1080/13854046.2017.1320426

27. Sherman, E., Brooks, BL., Iverson, GL., Slick, DL. and Strauss, E. (2011): Reliability and Validity in Neuropsychology. 10.1007/978-0-387-76978-3_30.

28. Chakrabartty, Satyendra Nath (2020): Reliability of test battery. *Methodological*

Innovations, 1–
8.10.1177/2059799120918340

29. Vaughan E D (1998): *Statistics: Tools for understanding data in the behavioral sciences*. Upper Saddle River, NJ: Prentice-Hall.

30. Chakrabartty, S, N. (2020 b): Improved Quality of Pain Measurement, *Health Science*, Vol. 1, 1 -6, 10.15342/hs.2020.259

31. Parkerson, HA., Noel, M., Gabrielle MP., Fuss, Samantha, Katz, J., Gordon J.G. Asmundson (2013): Factorial Validity of the English-Language Version of the Pain Catastrophizing Scale–Child Version, *The Journal of Pain,* 14 (11), 1383-1389, 10.1016/j.jpain.2013.06.004

32. Davies, D and Bouldin, W. (1979): A cluster separation measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1 (2), 224 – 227.

33. Chakrabartty, Satyendra Nath (2021): Integration of various scales for measurement of insomnia, *Research Methods in Medicine & Health Sciences*, 2(3), 102-111, 10.1177/26320843211010044.