

A Tool for Decision-Making in Norm-Referenced Survey Questionnaires with Items of Ordinal Variables

Ankur Barua^{1*}, Kumaraswamy Kademane¹, Biswadeep Das¹, Kumar Shiva Gubbiyappa²,
Rohit Kumar Verma², Sami Abdo Radman Al-Dubai¹

¹ School of Medicine, International Medical University (IMU), Malaysia

² School of Pharmacy, International Medical University (IMU), Malaysia

*** Corresponding Author:** Dr. Ankur Barua

Senior Lecturer, Department of Community Medicine,

International Medical University (IMU),

No. 126, Jalan Jalil Perkasa 19, Bukit Jalil, 57000 Kuala Lumpur, Malaysia

Email: ankurbarua26@yahoo.com

Mobile: +60122569902, +60105354023

Abstract

In order to arrive at a conclusion, setting up a cut-off point is necessary for opinion-based questionnaires on health care utilization, facilitating factors, barriers and also for assessing Knowledge, Attitude and Practice. This study has demonstrated on how to formulate a tool for decision-making in Norm-referenced survey questionnaires and readjust their cut-off points to incorporate the population variation for items containing ordinal variables. This procedure will help the researchers to perform finer adjustments in the cut-off values of any Norm-referenced survey instrument based on the local population data and in situations where no gold-standard instrument is available for comparison.

Key words: Point, cut-off, reliability, item, analysis, Cronbach's alpha, correlation

Background

Standardized testing involves the application of tests or instruments that are administered and scored in a pre-established standard or consistent manner.^{1,2} The quality or adequacy of any standardized testing instrument, whether norm-referenced or criterion-referenced, is directly associated with both reliability and validity studies.^{2,3} The use of standardized tests to conduct assessments is advantageous as the standardized tests yield quantifiable information, they can be used in screening programs. Standardized test results provide information regarding the participant's areas of strength and weakness. They can also be used to assess performance of an intervention method or disease progress over time.^{2,3,4} However, the most important advantage of results from a test administered in a standardized fashion is that it provides opportunities for drawing inference, generalization and extrapolation of findings to the whole community.^{2,4}

There are two types of standardized testing instruments namely “norm-referenced” tests and “criterion-referenced” tests.^{2,5} Academic achievement tests, cognitive impairment tests, intelligence quotient assessment (IQ) tests, well-being assessment tests are well known examples of norm-referenced, standardized tests. Norm-referenced test performance is generally summarized as one or more types of scores such as age-equivalence, grade-equivalence, percentile rankings, stanine, scaled scores, indexes, clusters or quotients.^{5,6} Newer editions of test instruments follow an item-response-theory procedure in their development which involves a new type of scoring system. This scoring system examines the difficulty level of each item in a questionnaire. The norm-referenced tests provide information on reliability and validity. They also provide language and presentation of items administration and scoring information as well as guidelines for the interpretation of the test results.^{7,8}

Methods

Item Analysis

In order to assess how well a test or an instrument is functioning we need to look at how well its individual items perform. Item analysis provides a way to exercise additional quality control over the tests by providing feedbacks on how successful the assessment actually was. An item analysis gets at the question of how well does it discriminate. If there are lots of items that didn't discriminate much at all then they need to be replaced by some better ones. Item analyses can also help the investigators diagnose why some items did not work especially well and suggest ways to improve them.^{3,4,5}

The characteristic features and behaviours of any two individuals in a community are never the same in every aspect. Every individual differs from one another at least in a minor dimension. Similarly, even in a standardized population, every community also varies from one another at least in a minor capacity. Though sometimes we consider things to be absolutely homogenous, but there always exist a minute amount of heterogeneity in every parameter of our assessment. Even if we achieve integrity in diversity of nature, we cannot really nullify the effect of diversity.^{2,3} So, it is not ideal to set up a universal cut-off in any survey instrument to arrive at a diagnosis from various communities with diverse socio-cultural backgrounds. It is also difficult to set up a cut-off point on overall items considered to assess Knowledge, Attitude and Practice (KAP) levels in KAP-based questionnaires to decide whether the overall knowledge of the respondents is adequate or not, their overall attitude is positive or negative and their overall practice is satisfactory or unsatisfactory. Setting up a cut-off point is also necessary for opinion-based questionnaires on health care utilization, facilitating factors and barriers to arrive at a conclusion of whether people are utilizing or recommending a procedure adequately or not, but it is often difficult to determine it. In this background, a study was conducted to formulate a tool for decision-making in Norm-referenced survey questionnaires and readjust their cut-off values to incorporate the population variation for items containing ordinal variables.

Determining the Cut-off Point of a Norm-referenced Test Instrument without any Gold-standard

This procedure is applicable for items having Ordinal Variables - e.g., scale of anxiety, scale of depression, scale of stress, scale of wellbeing where the respondents are instructed to provide their personal opinions or perceptions in an ordinal Likert scale. The weightage of each response in each item is directly proportional to the Discrimination Index (DI) as well as Internal Reliability or Cronbach's alpha. Hence, the weighted score for each response in each item is obtained by calculating the Observed Item Score multiplied by the product of Discrimination Index and Internal Reliability or Cronbach's alpha (Table 1). The "Correction Factor" is developed for making an adjustment in the overall cut-off value of the instrument. It is obtained from the ratio of the total weighted score and the total raw score. The overall cut-off value for the instrument is obtained by multiplying the "Correction Factor" with the 25th Percentile of each item and finally adding them up together (Table 1).

If the responses of the study population follow Normal Distribution:

The detailed mathematical model for the determination of a cut-off level for decision-making in Norm-referenced test instrument is described below:

(A) Calculation of Discrimination Index (DI) of individual items (refer Table 1)

= Point-biserial (Spearman's) Correlation Coefficient

(B) Weightage of each response in each item of the questionnaire (refer Table 1)

= (Observed Item Score) X (Discrimination Index) X (Internal Reliability or Cronbach's alpha)

(C) Correction Factor = [(Total Weighted score) / (Total Raw Score)] - Refer Table 2

(D) The cut-off point of an instrument without any gold standard (refer Table 2)

= Sum [(25th Percentile from Raw Score per Item) X (Correction Factor)]

A trial-run for this procedure is worked out on an anonymous, secondary database with five items in a Quality of Life (QOL) measuring scale rated in a six-point Likert Scale (Ordinal Scale) ranging from 0 to 5 per item (Tables 1 and 2). This instrument was earlier validated by the World Health Organization (WHO) and the cut-off set for QOL assessment was 13. The anonymous, secondary data source used was from a real-life, homogenous elderly population involving 609 respondents selected by simple random sampling technique to give it a test run on this new procedure. All the findings from the test run data were verified by a renowned psychiatrist for clinical confirmation by using the ICD-10 criteria. In this study, the Likert scale

was considered an Ordinal Scale for assessing Quality of Life (QOL) and data was analyzed in non-parametric scale due to the heterogeneous nature of the Likert Scale though there was normal distribution of a large and homogenous group of elderly respondents.

This method will work very well if the individual item correlation is between (0.75 and 1.0), ideal Cronbach's alpha is between (0.75 and 1.0) and also the ideal Correction Factor is between (0.75 and 1.0) for a database which is closer to normal distribution.

If the responses are collected from a Standardized Population:

A standardized population follows an ideal Normal Distribution (Gaussian) Curve where all the central tendencies like mean, median and mode coincide at the center and the dispersions like Standard Deviation and Interquartile Range are both equal to 1. If we assume that the normal database of a standardized population is broken into 100 equal parts then each individual unit is considered as a percentile. In that case, the Interquartile Range corresponds to 25th percentile and 75th percentile respectively. Suppose a participant scores the maximum possible marks (for example 100) allotted for a test; then in that database, 25th percentile will correspond to 25% and 75th percentile will correspond to 75% respectively. Considering the fact that everything is normal within 25th percentile and 75th percentile. Most of the common parameters in statistical methods are considered to be stronger if they are more than 75th Percentile or 75% in ideal situations. Hence, if the responses are collected from a Standardized Population then we need to consider the ideal individual item correlation as 0.75, ideal Cronbach's alpha as 0.75 and also the ideal Correction Factor as 0.75. Table 3 & 4 are examples of setting up an Ideal cut-off for a standardized population.

If the responses of the study population are skewed:

If the responses of the study population are skewed then Item Correlations, Cronbach's alpha as well as the Correction factor will be less than 0.75. In this condition, it is better to calculate the Median and its Interquartile Range. The 75th Percentile can be considered as the minimum cut-off for Positive Scoring Scales while 25th Percentile can be considered as the minimum cut-off for Negative Scoring Scales.

Results

Findings from the Test-run on Anonymous, Secondary Data Analysis

This QOL assessment instrument was earlier validated by the World Health Organization (WHO) and the cut-off set for decision on positive or negative levels of well-being status was set at 13. This was based on the crude mid-value of the minimum and maximum scores possible in this instrument. However, a test run with the anonymous, secondary data analysis using this new procedure revealed that the cut-off value should be readjusted to 11.3 for the population under study (Table 2). The interpretation follows that if the overall test score of any individual in this

QOL instrument becomes ≤ 11.3 then the person should be considered to be in a state of negative well-being status. It was found that there was a significant high level of agreement for well-being assessment at the two cut-off levels (Kappa value = 0.948 and $p=0.0001^*$). However, it was observed that if the cut-off level of this instrument is universally set at 13 then a significant percentage of individuals would be falsely labeled as of having negative well-being status by this instrument. This was also confirmed by the renowned psychiatrist who verified the data for clinical diagnosis by using the ICD-10 criteria. Since, this procedure is based on general statistical hypotheses; it can also be used to calculate a fresh cut-off value in a newly developed Norm-referenced questionnaire for obtaining diagnosis or conclusion purposes.

The findings reveal that this procedure can be used to perform finer adjustments in the cut-off values of any Norm-referenced survey instrument based on the local population data. This method can also be used for setting cut-off to arrive at a diagnosis in a newly developed instrument without any gold-standard instrument to compare it against. The Knowledge, Attitude and Practice (KAP) questionnaires are ideal for application of this procedure to help the researchers determine whether overall knowledge is adequate or not, overall attitude is positive or not and practice of the surveyed population is satisfactory or not. Cut-off can also be set for opinion-based questionnaires on health care utilization, facilitating factors and barriers to arrive at a conclusion of whether people are utilizing or recommending a procedure adequately or not. A pilot study with 10% of the estimated minimal sample size will help in identifying the cut-off value of the study instrument before undertaking the main study. Otherwise, the data from the main study can also be used for fixing the cut-off value before arriving at a diagnosis.

However, for knowledge and practice based questionnaires, the cut-off points need to be stratified according to the respective knowledge, attitude and practice sections. The above-mentioned procedure needs to be first applied on a standardized population assumed of having high level of knowledge and practice. The same procedure again needs to be applied on a standardized population assumed of having very low level of knowledge and practice. The average (Median) of both the two cut-off points from the Sum [(25th Percentile from Raw Score per Item) X (Correction Factor)] should then be considered as the actual cut-off points for the knowledge and practice sections separately. The assumptions on high and low levels of knowledge and practices can be drawn from previous studies or a descriptive study with focused group discussion prior to the main study. Bootstrapping can be done during data analysis when the composition and characteristics of a population is unknown and a study is conducted on the particular population for the first time.

Discussion

The index of discrimination is a useful measure of item quality. A basic consideration in evaluating the performance of a normative test item is the degree to which the item discriminates between responses which have high rates as compared to low rates. The Discrimination Index refers to how well an assessment differentiates between high and low scorers in an examination. We usually expect that the high-performing students would select the correct answer for each question more often than the low-performing students. If this is true, then the assessment is considered to have a positive discrimination index (between 0 and 1). This indicates that students

who had received a high total score chose the correct answer for a specific item more often than the students who had a lower overall score. If, however, we find that more of the low-performing students got a specific item correct, then the item has a negative discrimination index (between -1 and 0).^{2,9,10} The anonymous, secondary data analysis revealed a high Discrimination Index (DI) in terms of Spearman's Correlation Coefficient for the study instrument (Table 1).

Item discrimination indicates the extent to which success on an item corresponds to success on the whole test. Since all items in a test are intended to cooperate to generate an overall test score, any item with negative or zero discrimination undermines the test. The Discrimination Index (D) is computed from equal-sized high and low scoring groups on the test.^{9,10} The Point-biserial Correlation is the Spearman's correlation between responses to a particular item and scores on the total test (with or without that item).^{10,11,12}

Summated scales are often used in survey instruments to probe underlying constructs that the researcher wants to measure. These may consist of indexed responses to dichotomous or multi-point questionnaires, which are later summed to arrive at a resultant score associated with a particular respondent. The development of assessment scales also need to follow predictor variables for use in objective models. The concept of reliability rises as the function of scales is stretched to encompass the realm of prediction. Reliability tests are especially important when derivative variables are intended to be used for subsequent predictive analyses. If the scale shows poor reliability, then individual items within the scale must be re-examined and modified or completely changed as needed. One of the most popular reliability statistics in use today is Cronbach's alpha. It determines the internal consistency or average correlation of items in a survey instrument to gauge its internal reliability.^{12,13,14} The anonymous, secondary data analysis in this study revealed a high Cronbach's alpha value of 0.872 for the study instrument (Table 1). However, another good method of screening for efficient items is to run an exploratory factor analysis on all the items contained in the survey to weed out those variables that failed to show high correlation. No similar article was found related to this study as this was a new approach involving a new procedure to obtain standardized cut-off values in Norm-referenced test instruments or questionnaires for items containing ordinal variables.

Conclusion

There is a need to readjust and fine-tune the cut-off values for every survey instrument based on the population diversity before concluding any diagnosis. The procedure discussed in this study will help the researchers to perform finer adjustments in the cut-off values of any Norm-referenced survey instrument based on the local population data. This method can also be used for setting up a cut-off point to arrive at a diagnosis in a newly developed instrument which does not have any gold-standard instrument for comparison. A pilot study with 10% of the estimated minimal sample size will help in identifying the cut-off value of the study instrument before undertaking the main study. Otherwise, the data from the main study can also be used for fixing the cut-off value before arriving at a diagnosis. However, this method should not be used for any Criterion-referenced test or instrument.

Competing of Interest: There is no conflict of interest.

Author's Contribution

AB had conceived this project and developed the decision making tools. KK, BD and SAR gave it a test run on anonymous dataset and conducted secondary data analysis. RKV, KSG had compiled and finalized this report. All authors had read and approved this final draft.

Author's Information

Dr. ANKUR BARUA is working as a Senior Lecturer in the Department of Community Medicine at the International Medical University (IMU), Kuala Lumpur, Malaysia.

Ethical Considerations: No ethical issues involved as anonymous, secondary data analysis was done.

Acknowledgements

The author would like to extend his sincere and heartfelt gratitude to Dr. Dinker R. Pai, Professor & Head, Department of Surgery and Mr. Shanker Ram, Senior Lecturer, Department of Psychiatry at the Melaka-Manipal Medical College (MMMC), Melaka, Malaysia for providing their intellectual inputs and valuable guidance for this project.

References

1. American Psychological Association, National Council on Measurement in Education, & American Educational Research Association. Standards for educational and psychological testing. Washington, DC: American Psychological Association; 1999.
2. IRA/NCTE Joint Task Force on Assessment. Standards for the assessment of reading and writing. Newark, DE: International Reading Association and Urbana, IL: National Council of Teachers of English; 1994.
3. Lentz FE. Direct observation and measurement of academic skills: A conceptual review. In E. S. Shapiro & T. R. Kratochwill (Eds.). Behavioral assessment in schools. New York: Guilford; 1988: 76-120.
4. National Commission on Testing and Public Policy. From gatekeeper to gateway: Transforming testing in America. Chestnut Hill, MA: Boston College; 1990.

5. Quinto F, McKenna B. Alternatives to standardized testing. Washington, DC: National Education Association, Division of Instruction and Professional Development; 1977.
6. Mercer CD. Students with learning disabilities (5th ed.). Upper Saddle River, NJ: Prentice-Hall; 1997.
7. Weaver C. Understanding whole language: From principles to practice. Portsmouth, NH: Heinemann; 1990.
8. Witt JC, Elliot SN, Daly III EJ, Gresham FM, Kramer JJ. Assessment of at-risk and special needs children (2nd ed.). Boston: McGraw-Hill; 1988.
9. Fuchs LS, Fuchs D. Linking assessment to instructional intervention: An overview. *School Psychology Review* 1986; 15: 319–22.
10. Ebel RL. Problems of Communication between Test Specialists and Test Users. *Educational and Psychological Measurement* 1954; 14: 277-82.
11. Tzuriel D, Samuels MT. Dynamic assessment of learning potential: Inter-rater reliability of deficient cognitive functions, type of mediation, and non-intellective factors. *Journal of Cognitive Education and Psychology* 2000; 1: 41–64.
12. Symonds PM. Factors influencing test reliability. *Journal of Educational Psychology* 1928; 19: 73-87.
13. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297-334.
14. Santos JRA. Cronbach's Alpha: A Tool for Assessing the Reliability of Scales. *Journal of Extension* 1999; 37(2): 34-6.

Table 1: Calculation of Weighted Scores in a Quality of Life (QOL) scale with (0-5) Scores

Items	Discrimination Index (Spearman's Correlation Coefficient) (DI)	Cronbach's alpha (CA)	Sum Total of Individual weighted Scores = $\sum(x_i \times DI \times CA)$
Item 1	0.872	0.872	1471.34
Item 2	0.880	0.872	1516.30
Item 3	0.762	0.872	1510.33
Item 4	0.758	0.872	1284.94
Item 5	0.777	0.872	1565.13
Total Weighted Score	-	-	7348.04
Total Raw Score	Calculated previously		10416.00
Correction Factor (CF)	$= (\text{Total Weighted Score}) / (\text{Total raw Score})$ $= 7348.04 / 10416.00$ $= 0.705$		

Table 2: Calculation of Cut-off Value in a Quality of Life (QOL) Measurement Instrument with (0-5) Scores

Items	Median (50th Percentile)	25th Percentile	Correction Factor (CF)	(25th Percentile) X CF
Item 1	4	3	0.705	2.12
Item 2	4	3	0.705	2.12
Item 3	4	3	0.705	2.12
Item 4	4	3	0.705	2.12
Item 5	4	4	0.705	2.82
Score Range = (0-25) Adjusted Cut-off Value	-	-	-	11.3
Score Range = (0-25) Crude Mid-Value	-	-	-	13.0

Table 3: Calculation of Weighted Scores in a Quality of Life (QOL) scale with (0-5) Scores for a Standardized Population

Items	Ideal Discrimination Index (Spearman's Correlation Coefficient) (DI)	Ideal Cronbach's alpha (CA)	Standardized Sum Total of Individual weighted Scores = $\sum(x_i \times DI \times CA)$
Item 1	0.75	0.75	1875.00*
Item 2	0.75	0.75	1875.00*
Item 3	0.75	0.75	1875.00*
Item 4	0.75	0.75	1875.00*
Item 5	0.75	0.75	1875.00*
Total Weighted Score for 1000 respondents	-	-	9375.00
Total Raw Score for 1000 respondents	Calculated Previously (Considering mid-value for 1000 respondents)		12500.00
Ideal Correction Factor (CF)	$= (\text{Total Weighted Score}) / (\text{Total raw Score})$ $= 9375.00 / 12500.00$ $= 0.75$		

* Without this calculation also the Ideal Correction Factor can be assumed as 0.75.

Table 4: Calculation of Cut-off Value in a Quality of Life (QOL) Measurement Instrument with (0-5) Scores for a Standardized Population

Items	Median (standardized)	25th Percentile (standardized)	Ideal Correction Factor (CF)	(Standardized 25th Percentile) X CF
Item 1	2.5	2.45*	0.75	1.84
Item 2	2.5	2.45*	0.75	1.84
Item 3	2.5	2.45*	0.75	1.84
Item 4	2.5	2.45*	0.75	1.84
Item 5	2.5	2.45*	0.75	1.84
Score Range = (0-25) Adjusted Cut-off Value	-	-	-	9.2
Score Range = (0-25) Crude Mid-Value	-	-	-	13.0

* Considering Type1 (alpha) error as 0.05